Development of Novel Schemes for Treating Subsystem Boundaries and Electrostatic Potentials in Simulations of Complex Systems

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Bo Wang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Thesis advisor: Professor Donald G. Truhlar

January 2014

© Bo Wang 2014

Acknowledgements

I want to thank my advisor Professor Donald G. Truhlar for his guidance and support. He has greatly inspired me to move forward in my PhD research and has given me lots of supports and encouragement to get through the obstacles I met in the research. I also want to thank his group members, especially Miho Isegawa and Hannah Leverentz for helpful discussion and collaboration, and Masahiro Higashi, Hai Lin, Tao Yu, Yan Zhao, and Jingjing Zheng for their help and encouragement in my research.

I also want to thank Shuxia Zhang and other staff members in Minnesota Supercomputing Institute (MSI) for answering my questions about programing and the usage of supercomputer.

Finally I want to thank my parents and my wife for the love that supports me every day.

Dedication

This thesis is dedicated to my parents and my wife.

Abstract

Fragmentation schemes provide a powerful strategy for calculating the potential energy surfaces of complex systems. The combined quantum mechanical and molecular mechanical (QM/MM) method, the electrostatically embedded many-body (EE-MB) method, and the molecular tailoring approach (MTA) are three examples. Two critical issues to be addressed in these methods are the treatment of the boundary between the subsystems when it passes between bonded atoms and the inclusion of the electrostatic potential of one subsystem in the Hamiltonian of another. This thesis involves the development and application of new schemes to treat both issues. The first part focuses on the development of a tuned pseudoatom scheme with a balanced redistributed charge algorithm to accurately model the QM–MM boundary that passes through a covalent bond, especially a polar covalent bond. Various redistribution schemes and ways of tuning the boundary treatments are tested and compared for the QM/MM method and the EE-MTA method. The second part of this thesis involves the development of screened charge models to include charge penetration and screening effects in generating electrostatic potentials for use in various methods, including QM/MM and EE-MB methods. The screened charge models are also used to derive partial atomic charges by fitting electrostatic potentials.

Table of Contents

List of Tables

List of Figures

Preface

The following are citations for previously published works reproduced in this thesis:

Chapter 2.

Adapted with permission from Wang, B.; Truhlar, D. G., Combined Quantum Mechanical and Molecular Mechanical Methods for Calculating Potential Energy Surfaces: Tuned and Balanced Redistributed-Charge Algorithm, *J. Chem. Theory Comput.*, **2010**, *6*, 359–369. © 2010 American Chemical Society.

Chapter 3.

Adapted with permission from Wang, B.; Truhlar, D. G., Geometry Optimization Using Tuned and Balanced Redistributed Charge Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations, *Phys. Chem. Chem. Phys.*, **2011**, *13*, 10556–10564. http://pubs.rsc.org/en/content/articlelanding/2011/cp/c0cp02850a

Chapter 4.

Adapted with permission from Wang, B.; Truhlar, D. G., Tuned and Balanced Redistributed Charge Scheme for Combined Quantum Mechanical and Molecular Mechanical (QM/MM) Methods and Fragment Methods: Tuning Based on the CM5 Charge Model, *J. Chem. Theory Comput.*, **2013**, *9*, 1036–1042. © 2013 American Chemical Society.

Chapter 5.

Adapted with permission from Isegawa, M.; Wang, B.; and Truhlar, D. G., Electrostatically Embedded Molecular Tailoring Approach and Validation for Peptides, *J. Chem. Theory Comput.*, **2013**, *9*, 1381–1393. © 2013 American Chemical Society.

Chapter 6.

Adapted with permission from Wang, B.; Truhlar, D. G., Including Charge Penetration Effects in Molecular Modeling, *J. Chem. Theory Comput.*, **2010**, *6*, 3330–3342. © 2010 American Chemical Society.

Chapter 7.

Adapted with permission from Tempkin, J. O. B.; Leverentz, H. R.; Wang, B.; Truhlar, D. G., Screened Electrostatically Embedded Many-Body Method, *J. Phys. Chem. Lett.*, **2011**, *2*, 2141–2144. © 2011 American Chemical Society.

Chapter 8.

Adapted with permission from Wang, B.; Truhlar, D. G., Partial Atomic Charges and Screened Charge Models of the Electrostatic Potential, *J. Chem. Theory Comput.*, **2012**, *8*, 1989–1998. © 2012 American Chemical Society.

Chapter 1. Introduction

1.1 General overview

Computational simulations provide an important tool to understand physical and chemical changes at the atomic level.[1, 2] Two classes of methods have been widely applied to construct the potential energy surfaces for molecules and solids. One class involves quantum mechanical electronic structure calculations, in which the Schrödinger equation[3] or the Kohn-Sham equations[4] are solved for the molecules or solids. The other class involves classical mechanical force field methods, in which a force field is parameterized against experimental values or quantum mechanical calculations for certain kinds of molecules, for example biomolecules.[5, 6]

In the study of complex systems such as zeolites, biopolymers, or large clusters, it is usually too expensive to perform full quantum mechanical calculations, and force fields may be inadequate if one needs to accurately describe chemical reactions in the system. One widely used strategy is partitioning the system into subsystems or fragments. The total energy of the system is then derived from a linear combination of the energies of subsystems and the interaction energies among the subsystems.[7] The combined quantum mechanical/molecular mechanical (QM/MM) method is a special case in which the system is partitioned into two regions – a small region is treated by quantum mechanics and the remaining part treated by molecular mechanics.[8-23]

Other fragmentation methods include the molecular tailoring approach (MTA),[24-32] the fragment molecular orbital (FMO) method,[33-41] the molecular fractionation with

conjugate caps (MFCC) method,[42, 43] the effective fragment potential (EFP),[44] the systematic molecular fragmentation (SMF) method,[45-47] the generalized energy-based fragmentation (GEBF) method,[48-50] the explicit polarization (X-Pol) method,[51, 52] the kernel energy method (KEM),[53, 54] the electrostatically embedded many-body (EE-MB) method,[55-58] the ternary interaction model,[59] the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE),[56, 57] the molecules in molecules (MIM) method,[60] the multilevel fragment-based approach (MFBA),[61] the hybrid many-body interaction (HMBI) method,[62] and the many-overlapping-body expansion.[63]

In various fragmentation methods, the quality of the results depends critically on the treatment of the subsystems and their interactions. To better incorporate the effects of the surroundings on the subsystem, different levels of approximation have been proposed.[7, 21, 64] One of the most effective approaches is to embed the subsystem in the electrostatic potentials of other subsystems, as is done in the electrostatically embedded QM/MM method[21] and the EE-MB method.[55-58] This kind of treatment allows the polarization of the subsystems by the surroundings and greatly improves the performance of the methods as compared to the performance achieved without embedding. Point charges centered on atoms are widely applied to represent the electrostatic potentials in various fragmentation methods.[7, 21] Charge densities or effective embedding potentials, such as *ab initio* model potentials (AIMP),[65, 66] are also used in more sophisticated embedding schemes. Moreover, in the systems that contain boundaries between the subsystems through covalent bonds, special treatments of subsystems are needed to alleviate the errors arising from the differences between the subsystems and the original system.[21]

In the present thesis, we mainly concentrate on two issues. One is the treatment of the boundary between subsystems when that boundary crosses covalent bonds, such as some QM-MM boundaries in the QM/MM method; and the other is the treatment of electrostatic interactions between a subsystem and its surroundings.

1.2 Boundary treatment

In the QM/MM method and fragmentation methods, the treatment of the boundary between two regions is an important issue. This is especially difficult if the boundary passes through a bond, which is practically unavoidable in the treatment of many solids, polymers, and complex systems. In general, the region treated by quantum mechanics is capped to saturate dangling valences caused by the cut. Here we summarize the methods used in various QM/MM methods to deal with capping the quantum mechanical (QM) boundary atom. Three different kinds of methods have been proposed. The first one is the link atom (LA) approach.[9, 11] The dangling bond of the QM region is capped with an additional atom (usually a hydrogen atom), and the QM calculations are performed on this capped system. The second method is the localized orbitals approach.[67] The dangling bond is saturated by orbitals rather than by an atom. Examples of this approach are the local self-consistent field (LSCF) method[67] and the generalized hybrid orbitals (GHO) scheme.[68, 69] The third kind of method involves a pseudobond or an effective core potential (ECP). In this approach, a parameterized atom, modified to mimic the behavior of the original MM boundary atoms or groups, is used to cap the QM system; examples of this approach are tuned capping atoms,[70-72] adjusted connection atoms,[73] a pseudobond,[74-76] an effective group potential,[77] a quantum capped potential,[78-80] and a variationally optimized effective atom-centered potential.[81] This third class of methods

may be considered to be a second-generation link-atom method in which the link atom is optimized or tuned.

Another issue to consider for the boundary is the treatment of partial atomic charges near the boundary. It has been found that it is important to conserve the total charge of the entire QM/MM system in QM/MM calculations,[82] that is, the sum of the MM partial atomic charges of the MM region and the QM charge of the capped QM region should equal the total charge of the original entire system, as shown in eq 1.1:

$$q^{\mathrm{MM}} + q^{\mathrm{QM}} = q^{\mathrm{total}} \tag{1.1}$$

When the partial atomic charges satify eq 1.1, we say that they are balanced. However, when the original entire system is divided into QM and MM regions, the sum of MM charges of the MM region does not necessarily equal zero or an integer. If MM charges are not modified, the total charge of the QM/MM entire system is not conserved. Several workers have recognized that this causes inaccuracies and have suggested various methods to remedy this problem.[74, 82-84]

In the present thesis, we develop tuned and balanced charge schemes to better treat the boundary. The quantum mechanical region is properly terminated by a tuned pseudopotential to mimic the original bond, and the surrounding partial atomic charges are properly balanced and redistributed to conserve the total charge of the systems and to avoid overpolarization of the quantum region.

1.3 Electrostatic interactions between subsystems

Electrostatic interactions between molecules and subsystems can be represented by eq 1.2:

$$E^{\mathrm{el}} = \iint \frac{\rho_A^{\mathrm{total}}(\vec{r}_1)\rho_B^{\mathrm{total}}(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d^3\vec{r}_1 d^3\vec{r}_2 \tag{1.2}$$

where $\rho_A^{\mathrm{total}}(\vec{r}_1)$ and $\rho_B^{\mathrm{total}}(\vec{r}_2)$ are the total charge densities of the molecules or subsystems *A* and *B*. This term usually dominates the interactions between subsystems. To simplify the evaluation, most modelers make the distributed monopole approximation, in which the total density of the molecule is approximated as the sum of point charges centered on atoms. In most methods and compuer programs for complex-system simulations, the coulomb interactions between these partial atomic charges play an important role in computing the electrostatic interactions.

In the electrostatically embedded calculations, the subsystem is embedded in the electrostatic potential of the surroundings. The electrostatic interactions between subsystems are evaluated by the interactions between the quantum mechanical (QM) density of one subsystem and the electrostatic potential of the surroundings. In the monopole approximation, the electrostatic potential of the surroundings is represented by a set of partial atomic charges, which are point charges centered on atoms. These point charges are added as one-electron integrals into the QM Hamiltonian, as shown in eq 1.3, allowing the polarization of the QM electron density by the environment.[11,15]

$$H_{\mathrm{QM/MM}}^{\mathrm{el}} = -\sum_{i,A} \frac{q_A}{r_{iA}} + \sum_{\alpha,A} \frac{Z_\alpha q_A}{R_{\alpha A}} \tag{1.3}$$

where q_A are the point charges of the surrounded atoms; the indices *i* and α run over all QM electrons and nuclei, respectively; and r_{iA} and $R_{\alpha A}$ are the distances between the QM electrons and the surrounded point charges and the distances between the QM nuclei and the surrounded point charges, respectively.

However, a point charge model is not necessarily a good model for the charge density of the molecules and the system. Four possible improvements can be considered, in particular the addition of higher-order multipole contributions at each nuclear center,[85-87] the use of off-nuclei point charges,[88, 89] the inclusion of penetration effects,[90-92] and the treatment of additional quantum mechanical effects associated with the distributed charge distribution.[93-105] In a general way, the first three approaches all account for the same effect, namely that the actual electron density has more structure than a collection of point charges. The distributed multipole method accounts for the asymmetry of the charge distribution of an atom in a molecule, the use of off-nuclei charge centers accounts for both asymmetry and finite orbital extent, and the penetration modeling accounts for finite orbital extent. One can also include both higher multipole moments and penetration effects.[98, 106-108]

In the present thesis, we focus on the penetration effects. The essence of penetration effects, which cannot be described by MM point charges or distributed multipoles, is that when two atoms are close enough, their charge densities can overlap, which decreases the shielding of the nuclear charge of each atom by its own electron density. Various approaches have been suggested to include this effect in both MM calculations and QM/MM calculations.[91, 92, 94, 98, 106-119] In this thesis, we have developed new screened charge models to be used in QM/MM methods and various fragmentation methods.

1.4 Organization of the thesis

This thesis is organized to address two issues. One is the development of the new boundary treatments in the combined quantum mechanical and molecular mechanical (QM/MM) methods and electrostatically embedded molecular tailoring approach (EE-

MTA), shown in Chapters 2, 3, 4 and 5. Tuned and balanced redistributed charge schemes with and without charge smearing scheme are developed to better treat covalent bonds between the QM and MM regions in these chapters. The other issue is the development of screened charge models used in the molecular mechanics (MM), the electrostatically embedded-many body (EE-MB) method, and the QM/MM methods, shown in Chapters 6, 7, and 8. Two new screened charge models are proposed to include charge penetration effects and describe the electrostatic potentials more accurately, and have been applied in various ways.

Chapter 2 describes the development of a new algorithm to treat the boundary across a covalent polar bond in the QM/MM method. The MM point charge on the MM boundary atom is modified ("balanced") to conserve the total charge of the entire system, and the modified charge is redistributed to the midpoints of the bonds between an MM boundary atom and its neighboring MM atoms. A tuned link atom is used to terminate the dangling bond of the QM region and to reproduce the partial charge of the uncapped portion of the QM subsystem. The method is applied to calculate proton affinities with various types of bonds being cut.

Chapter 3 of this thesis investigates the geometry optimization using the tuned and balanced redistributed charge schemes developed in Chapter 2. The redistributed charges near the QM–MM boundary are smeared to make the electrostatic interactions between the QM region and the redistributed charges more realistic.

Chapter 4 demonstrates a new way to carry out the tuning process; in particular, the CM5 charge model, rather than the Mulliken population analysis applied in previous studies, is used for tuning the capping atom that terminates the dangling bond of the QM

region. Compared with the Mulliken charges, the CM5 charges better describe the charge distributions in test molecules, and they are less dependent on the choice of basis set.

Chapter 5 applies the tuned and balanced redistributed-charge scheme into the study of electrostatically embedded molecular tailoring approach (EE-MTA) methods. The boundary is treated by generically tuned link atoms. These generically tuned cap atoms show better performance than the hydrogen cap atom for both the electronic energy and the energy difference between an α helix and a β sheet in EE-MTA methods.

Chapter 6 proposes a new scheme to include charge penetration effects in electrostatic modeling. In the new scheme, the MM atomic charge density of an atom in a molecule is represented by a screened charge rather than by a point charge. The screened charge includes a point charge for the nucleus, core electrons, and inner valence electrons, and a smeared charge for the outer valence electron density, which is distributed in a Slater-type orbital representing the outer part of the atomic charge distribution such that the resulting pairwise interactions are still analytic central potentials. The new model is parametrized and illustrated in the electronically embedded combined quantum mechanical and molecular mechanical (QM/MM) method. The test suite contains 40 dimers. Symmetry-adapted perturbation theory (SAPT)[120] is applied as the benchmark.

Chapter 7 of this thesis applies the screened charge model to the electrostatically embedded many body (EE-MB) method. Screened charges, instead of point charges, are used to represent the electrostatic potentials in the EE-MB method. The test suite used for evaluation is the binding energy for five water hexamers.

Chapter 8 demonstrates that screened charge models can be also used to derive the partial atomic charges by fitting the electrostatic potentials. The quality of the fit to the

electrostatics is improved in the screened charge methods, especially for the regions that are within one van der Waals radius of the centers of atoms. It is also found that the charges derived by fitting electrostatic potentials with screened charges are less sensitive to the positions of the fitting points than those derived with the point charge model.

1.5 References for Chapter 1

1. C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons Inc, 2004.
2. R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, New York, 2004.
3. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry-Introduction to Advanced Electronic Structure Theory*, Dover Publications, Inc., Mineola, New York, 1989.
4. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133-A1138.
5. D. Alexander and J. Mackerell, *J. Comp. Chem.*, 2004, **25**, 1584-1604.
6. J. W. Ponder and D. A. Case, *Adv.Protein Chem.*, 2003, **66**, 27-85.
7. M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chemical Reviews*, 2012, **112**, 632-672.
8. A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227-249.
9. U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, 1986, **7**, 718-730.
10. J. Gao, *Rev. Comp. Chem*, 1996, **7**, 119-185.
11. D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580-10594.
12. I. Antes and W. Thiel, *J. Phys. Chem. A*, 1999, **103**, 9290-9295.
13. N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720-1735.
14. P. Sherwood, in *Hybrid Quantum Mechanics/Molecular Mechanics Approaches*, ed. J. Grotendorst, John von Neumann Institute for Computing, Jülich, 2000, vol. 3, p. 285.
15. J. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467-505.
16. A. Shurki and A. Warshel, in *Adv.Protein Chem.*, Academic Press Inc, San Diego, USA, 2003, vol. 66, pp. 249-313.
17. R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389-427.
18. M.-E. Moret, E. Tapavicza, L. Guidoni, U. Röhrig, M. Sulpizi, I. Tavernelli and U. Rothlisberger, *Chimia*, 2005, **59**, 493-498.
19. T. Vreven, K. Morokuma and C. S. David, *Annual Reports in Computational Chemistry*, 2006, **2**, 35-51.
20. H. M. Senn and W. Thiel, *Curr. Opin. Chem. Biol.*, 2007, **11**, 182-187.
21. H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185-199.
22. H. Hu and W. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573-601.
23. H. M. Senn and W. Thiel, *Angew. Chemie Int. Ed.*, 2009, **48**, 1198-1229.
24. S. R. Gadre, R. N. Shirsat and A. C. Limaye, *J. Phys. Chem.*, 1994, **98**, 9165-9169.
25. K. Babu and S. R. Gadre, *J. Comp. Chem.*, 2003, **24**, 484-495.

26. V. Ganesh, R. K. Dongare, P. Balanarayan and S. R. Gadre, *J. Chem. Phys.*, 2006, **125**, 104109.
27. M. Elango, V. Subramanian, A. P. Rahalkar, S. R. Gadre and N. Sathyamurthy, *J. Phys. Chem. A*, 2008, **112**, 7699-7704.
28. A. P. Rahalkar, M. Katouda, S. R. Gadre and S. Nagase, *J. Comp. Chem.*, 2010, **31**, 2405-2418.
29. A. P. Rahalkar, V. Ganesh and S. R. Gadre, *J. Chem. Phys.*, 2008, **129**, 234101.
30. S. D. Yeole and S. R. Gadre, *J. Chem. Phys.*, 2010, **132**, 094102.
31. A. S. Mahadevi, A. P. Rahalkar, S. R. Gadre and G. N. Sastry, *J. Chem. Phys.*, 2010, **133**, 164308.
32. S. D. Yeole and S. R. Gadre, *J. Chem. Phys.*, 2011, **134**, 084111.
33. K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **313**, 701-706.
34. T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi and K. Kitaura, *Chem. Phys. Lett.*, 2000, **318**, 614-618.
35. K. Kitaura, S. I. Sugiki, T. Nakano, Y. Komeiji and M. Uebayasi, *Chem. Phys. Lett.*, 2001, **336**, 163-170.
36. D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2004, **120**, 6832-6840.
37. D. G. Fedorov, T. Ishida and K. Kitaura, *J. Phys. Chem. A*, 2005, **109**, 2638-2646.
38. D. G. Fedorov, K. Ishimura, T. Ishida, K. Kitaura, P. Pulay and S. Nagase, *J. Comp. Chem.*, 2007, **28**, 1476-1484.
39. T. Sawada, D. G. Fedorov and K. Kitaura, *Int. J. Quantum Chem.*, 2009, **109**, 2033-2045.
40. K. Murata, D. G. Fedorov, I. Nakanishi and K. Kitaura, *J. Phys. Chem. B*, 2009, **113**, 809-817.
41. D. G. Fedorov, J. H. Jensen, R. C. Deka and K. Kitaura, *J. Phys. Chem. A*, 2008, **112**, 11808-11816.
42. D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599-3605.
43. N. Jiang, J. Ma and Y. S. Jiang, *J. Chem. Phys.*, 2006, **124**, 114112.
44. M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys and W. J. Stevens, *J. Phys. Chem. A*, 2001, **105**, 293-307.
45. V. Deev and M. A. Collins, *J. Chem. Phys.*, 2005, **122**, 154102.
46. M. A. Collins and V. A. Deev, *J. Chem. Phys.*, 2006, **125**, 104104.
47. S. R. Pruitt, M. A. Addicoat, M. A. Collins and M. S. Gordon, *Phys. Chem. Chem. Phys*, 2012, **14**, 7752-7764.
48. W. Li, S. Li and Y. Jiang, *J. Phys. Chem. A*, 2007, **111**, 2193-2199.
49. W. Hua, T. Fang, W. Li, J.-G. Yu and S. Li, *J. Phys. Chem. A*, 2008, **112**, 10864-10872.
50. S. Hua, W. Hua and S. Li, *J. Phys. Chem. A*, 2010, **114**, 8126-8134.
51. W. Xie, L. Song, D. G. Truhlar and J. Gao, *J. Chem. Phys.*, 2008, **128**, 234108.
52. Y. Wang, C. P. Sosa, A. Cembran, D. G. Truhlar and J. Gao, *J. Phys. Chem. B*, 2012, **116**, 6781-6788.
53. L. Huang, L. Massa and J. Karle, *Int. J. Quantum Chem.*, 2005, **103**, 808-817.
54. L. Huang, L. Massa and J. Karle, *Int. J. Quantum Chem.*, 2006, **106**, 447-457.
55. E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 46-53.
56. E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 1342-1348.

57. E. E. Dahlke, H. R. Leverentz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2008, **4**, 33-41.
58. H. R. Leverentz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2009, **5**, 1573-1584.
59. S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki and H. Sekino, *Molecular Physics*, 2005, **103**, 2255-2265.
60. N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.*, 2011, **7**, 1336-1343.
61. J. Řezač and D. R. Salahub, *J. Chem. Theory Comput.*, 2009, **6**, 91-99.
62. S. Wen, K. Nanda, Y. Huang and G. J. O. Beran, *Phys. Chem. Chem. Phys*, 2012, **14**, 7578-7590.
63. N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.*, 2012, **8**, 2669-2675.
64. J. Weber and J. Gunne, *Phys. Chem. Chem. Phys*, 2010, **12**, 583-603.
65. Z. Barandiarán and L. Seijo, *J. Chem. Phys.*, 1988, **89**, 5739-5746.
66. M. A. Nygren, L. G. M. Pettersson, Z. Barandiarán and L. Seijo, *J. Chem. Phys.*, 1994, **100**, 2010-2018.
67. V. Théry, D. Rinaldi, J. L. Rivail, B. Maigret and G. G. Ferenczy, *J. Comp. Chem.*, 1994, **15**, 269-282.
68. J. Gao, P. Amara, C. Alhambra and M. J. Field, *J. Phys. Chem. A*, 1998, **102**, 4714-4721.
69. J. Pu, J. Gao and D. G. Truhlar, *J. Phys. Chem. A*, 2004, **108**, 632-650.
70. A. Redondo, W. A. Goddard, C. A. Swarts and T. C. McGill, *J. Vac. Sci. Technol.*, 1981, **19**, 498-501.
71. N. Koga and K. Morokuma, *Chem. Phys. Lett.*, 1990, **172**, 243-248.
72. V. A. Nasluzov, E. A. Ivanova, A. M. Shor, G. N. Vayssilov, U. Birkenheuer and N. Rösch, *J. Phys. Chem. B*, 2003, **107**, 2228-2241.
73. I. Antes and W. Thiel, *J. Phys. Chem. A*, 1999, **103**, 9290-9295.
74. Y. Zhang, T.-S. Lee and W. Yang, *J. Chem. Phys.*, 1999, **110**, 46-54.
75. Y. Zhang, *J. Chem. Phys.*, 2005, **122**, 024114.
76. J. M. Parks, H. Hu, A. J. Cohen and W. Yang, *J. Chem. Phys.*, 2008, **129**, 154106.
77. F. Alary, R. Poteau, J. L. Heully, J. C. Barthelat and J. P. Daudey, *Theor. Chem. Acc.*, 2000, **104**, 174-178.
78. G. A. DiLabio, M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.*, 2002, **116**, 9578-9584.
79. S. Moon, P. A. Christiansen and G. A. DiLabio, *J. Chem. Phys.*, 2004, **120**, 9080-9086.
80. G. A. DiLabio, R. A. Wolkow and E. R. Johnson, *J. Chem. Phys.*, 2005, **122**, 044708.
81. O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *J. Chem. Phys.*, 2005, **122**, 014113.
82. R. C. Walker, M. F. Crowley and D. A. Case, *J. Comp. Chem.*, 2008, **29**, 1019-1031.
83. P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent and I. H. Hillier, *Faraday Discuss.*, 1997, **106**, 79-92.
84. D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček and B. R. Brooks, *J. Chem. Phys.*, 2002, **117**, 10534-10547.
85. A. J. Stone, *Chem. Phys. Lett.*, 1981, **83**, 233-239.
86. W. A. Sokalski and R. A. Poirier, *Chem. Phys. Lett.*, 1983, **98**, 86-92.

87. H. Leverentz, J. Gao and D. G. Truhlar, *Theor. Chem. Acc.*, in press.
88. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
89. G. C. Hancock, D. G. Truhlar and C. E. Dykstra, *J. Chem. Phys.*, 1988, **88**, 1786-1796.
90. A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press Inc., New York, 1996.
91. D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček and B. R. Brooks, *J. Chem. Phys.*, 2002, **117**, 10534-10547.
92. G. A. Cisneros, S. N.-I. Tholander, O. Parisel, T. A. Darden, D. Elking, L. Perera and J. P. Piquemal, *Int. J. Quantum Chem.*, 2008, **108**, 1905-1912.
93. T. Bredow, G. Geudtner and K. Jug, *J. Chem. Phys.*, 1996, **105**, 6395-6400.
94. P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch and D. Cohen, *J. Chem. Phys.*, 1996, **105**, 1968-1986.
95. I. V. Yudanov, V. A. Nasluzov, K. M. Neyman and N. Rösch, *Int. J. Quantum Chem.*, 1997, **65**, 975-986.
96. N. López and F. Illas, *J. Phys. Chem. B*, 1998, **102**, 1430-1436.
97. G. Pacchioni and A. M. Ferrari, *Catal. Today*, 1999, **50**, 533-540.
98. M. A. Freitag, M. S. Gordon, J. H. Jensen and W. J. Stevens, *J. Chem. Phys.*, 2000, **112**, 7300-7306.
99. R. Soave and G. Pacchioni, *Chem. Phys. Lett.*, 2000, **320**, 345-351.
100. J. R. B. Gomes, F. Illas, N. C. Hernández, A. Márquez and J. F. Sanz, *Phys. Rev. B*, 2002, **65**, 125414.
101. J. R. B. Gomes, F. Illas, N. C. Hernández, J. F. Sanz, A. Wander and N. M. Harrison, *J. Chem. Phys.*, 2002, **116**, 1684-1691.
102. J. R. B. Gomes, Z. Lodziana and F. Illas, *J. Phys. Chem. B*, 2003, **107**, 6411-6424.
103. F. Cinquini, C. D. Valentin, E. Finazzi, L. Giordano and G. Pacchioni, *Theor. Chem. Acc.*, 2007, **117**, 827-845.
104. R. Valero, J. R. B. Gomes, D. G. Truhlar and F. Illas, *J. Chem. Phys.*, 2008, **129**, 124710-124717.
105. A. Cembran, P. Bao, Y. Wang, L. Song, D. G. Truhlar and J. Gao, *J. Chem. Theory Comput.*, 2010, **6**, 2469-2476.
106. A. M. Köster, C. Kölle and K. Jug, *J. Chem. Phys.*, 1993, **99**, 1224-1229.
107. V. Kairys and J. H. Jensen, *Chem. Phys. Lett.*, 1999, **315**, 140-144.
108. J. P. Piquemal, N. Gresh and C. Giessner-Prettre, *J. Phys. Chem. A*, 2003, **107**, 10353-10359.
109. G. G. Hall and C. M. Smith, *Int. J. Quantum Chem.*, 1992, **42**, 1237-1250.
110. R. J. Wheatley and J. B. O. Mitchell, *J. Comput. Chem.*, 1994, **15**, 1187-1198.
111. B. Guillot and Y. Guissani, *J. Chem. Phys.*, 2001, **114**, 6720-6733.
112. P. Amara and M. J. Field, *Theor. Chem. Acc.*, 2003, **109**, 43-52.
113. J. P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh and T. A. Darden, *J. Chem. Phys.*, 2006, **124**, 104101.
114. M. Torheyden and G. Jansen, *Mol. Phys.*, 2006, **104**, 2101-2138.
115. M. A. Spackman, *Chem. Phys. Lett.*, 2006, **418**, 158-162.
116. G. A. Cisneros, J. P. Piquemal and T. A. Darden, *J. Phys. Chem. B*, 2006, **110**, 13682-13684.

117. L. V. Slipchenko and M. S. Gordon, *J. Comput. Chem.*, 2007, **28**, 276-291.
118. D. M. Elking, G. A. Cisneros, J. P. Piquemal, T. A. Darden and L. G. Pedersen, *J. Chem. Theory Comput.*, 2010, **6**, 190-202.
119. R. Kumar, F.-F. Wang, G. R. Jenness and K. D. Jordan, *J. Chem. Phys.*, 2010, **132**, 014309-014312.
120. B. Jeziorski, R. Moszynski and K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887-1930.

Chapter 2. Combined Quantum Mechanical and Molecular Mechanical Methods for Calculating Potential Energy Surfaces: Tuned and Balanced Redistributed-Charge Algorithm[1]

2.1 Introduction

The application of quantum chemistry to large and complex systems is one of the most challenging areas of current computational chemistry and also one that is seeing the most progress.[1] An important tool for such applications is the combined quantum mechanical/molecular mechanical (QM/MM) method for calculating potential energy surfaces and interatomic forces; the reader is directed to several reviews and overviews for background information.[2-23]

A stubborn issue in QM/MM calculations is the treatment of the boundary between QM and MM regions when it passes through a bond, which is practically unavoidable in the treatment of many solids, polymers, and complex systems. In general the QM region is capped to saturate dangling valences caused by the cut. Three different kinds of methods have been proposed to deal with capping the QM boundary atom. The first one is the link atom approach (LA).[24,25] The dangling bond of the QM region is capped with an additional atom (usually a hydrogen atom) and the QM calculations are performed on

[1] The authors thank Masahiro Higashi, Hai Lin, Manjeera Mantina, and Jingjing Zheng for helpful discussions. This work was supported in part by the National Science Foundation under grant no. CHE07-04974 and by the Office of Naval Research under Award Number N00014-05-1-0538.

this capped system. The second method is localized orbitals.[26-28] The dangling bond is saturated by orbitals rather than by an atom. Examples of this approach are the local self-consistent field (LSCF) method[26] and the generalized hybrid orbitals (GHO) scheme.[27,28] The third kind of method involves a pseudobond or an effective core potential (ECP). In this approach, a parameterized atom, modified to mimic the behavior of the original MM boundary atoms or groups, is used to cap the QM system; examples of this approach are tuned capping atoms,[29-31] adjusted connection atoms,[32] a pseudobond,[33-35] an effective group potential,[36] a quantum capped potential,[37-39] and a variationally optimized effective atom-centered potential.[40] This third class of methods may be considered to be a second-generation link-atom method in which the link atom is optimized or tuned.

Though much progress has been made, there are still many problems in the treatment of QM-MM boundaries that pass through a bond. Most attention has been devoted to the cutting of C–C bonds, especially for modeling enzymatic binding and reactions, but some procedures are more general. The methods that have been developed exhibit a wide variety of differences in the precise way in which they have been implemented.

Pople has emphasized the importance of theoretical models, where a theoretical model is "an approximate but well defined mathematical procedure for simulation…. The approximate mathematical treatment must be precisely formulated. It should be general…. Particular procedures for particular molecules…should be avoided."[41] If tests of the model against a broad data set are successful, the model is said to be validated. The goal of this article is to develop and validate a new method, in the spirit of a theoretical model chemistry, for the treatment of a boundary between bonded atoms in QM/MM simulations. It is precisely defined in a general way applicable to all systems and all kinds

of single bonds, and it is tested against a data set of 25 systems in which 13 different kinds of bonds are cut, in particular (where the atom listed first is in the QM subsystem, and the one listed second is in the MM subsystem): C–C, N–C, O–C, S–C, C–N, O–N, C–O, Al–O, Si–O, C–Si, O–Si, C–S, and S–S.

2.2 Method

Our group has developed redistributed charge (RC) and redistributed charge and dipole (RCD) methods to treat the charges near a QM-MM boundary that passes through a bond.[42] These methods give good results even when large charges are present near the boundary. In the current work, we improve the RC and RCD methods by adding two new elements – a charge balancing step and a tuned link atom. In particular, the redistributed charges are used to conserve the charge of the entire system, and a tuned fluorine atom is used to saturate the free valence of the QM region and to reproduce the partial charge of the uncapped portion of the capped QM subsystem. The improved method is used to treat polar bonds between the QM and MM subsystems with large partial atomic charges near the boundary. In order to describe the algorithm, we label the atoms according to "tiers". The definition is the same as what is used in previous work;[4,42] in particular, the MM boundary atoms are denoted as M1 atoms; and the MM atoms directly bonded to M1 atoms are denoted as M2 atoms. M3 atoms are the third–tier MM atoms. The QM boundary atoms are denoted as Q1 atoms. The QM atoms directly bonded to Q1 atoms are labeled Q2 atoms. Q3 atoms are those bonded to Q2 atoms, and so forth for Q4, Q5, etc. The QM region is also called the primary subsystem (PS) in this study. The sum of all QM atoms and MM atoms before the cutting and capping is called the original entire

system. The sum of the capped QM subsystem and the whole MM subsystem after the charge redistribution is called the QM/MM entire system.

In the QM/MM calculations, we use an additive QM/MM scheme to define the total energy of the system:[23]

$$E = E^{\mathrm{QM}} + E^{\mathrm{QM/MM}} + E^{\mathrm{MM}} \tag{2.1}$$

$$E^{\mathrm{QM/MM}} = E_{el}^{\mathrm{QM/MM}} + E_{vdW}^{\mathrm{QM/MM}} + E_{val}^{\mathrm{QM/MM}} \tag{2.2}$$

where E^{QM} is the quantum mechanical energy of the QM region, E^{MM} is the molecular mechanical energy of the MM region, and $E^{\mathrm{QM/MM}}$ accounts for the interaction energy between the QM and the MM regions. $E^{\mathrm{QM/MM}}$ is decomposed into three terms; $E_{el}^{\mathrm{QM/MM}}$ represents electrostatic interactions, $E_{vdW}^{\mathrm{QM/MM}}$ represents van der Waals interactions, and $E_{val}^{\mathrm{QM/MM}}$ represents valence interactions. In this study, we will concentrate on the electrostatic coupling term $E_{el}^{\mathrm{QM/MM}}$, which is the most technically involved term. The $E_{vdW}^{\mathrm{QM/MM}}$ and $E_{val}^{\mathrm{QM/MM}}$ terms will cancel out in the present work because we study fixed-geometry proton affinities to isolate the electrostatic terms, but these other QM/MM terms will be studied later when we consider QM/MM geometry optimization.

2.2.1 Treatment of boundary charge

It has been found that it is important to conserve the total charge of the QM/MM entire system in QM/MM calculations,[43] that is, the sum of the MM partial atomic charges of the MM region and the QM charge of the capped QM region should equal the total charge of the original entire system, as shown in eq 2.3:

$$q^{\mathrm{MM}} + q^{\mathrm{QM}} = q^{\mathrm{total}} \tag{2.3}$$

However, when the original entire system is divided into QM and MM regions, the sum of MM charges of the MM region does not necessarily equal zero or an integer. If MM charges are not modified, the total charge of the QM/MM entire system is not conserved. Several workers have recognized that this causes inaccuracies and have suggested various methods to remedy this.[33,43-45] Sherwood et al.[44] adjusted the charge on an M1 atom to conserve the total charge of the QM/MM entire system, and they redistributed the adjusted charge on the M1 atom evenly to M2 atoms; point dipoles were added at the M2 atoms to compensate the changes in the M1–M2 bond dipoles due to the movement of the charges. Zhang et al.[33] zeroed the charges on all MM atoms that are in the same group as the M1 atom. Das et al.[45] used a double link atom approach combined with delocalized Gaussian MM charges. Walker et al.[43] added the charge difference to the nearest M2 atom or evenly to all the MM atoms except the M1 atom.

In the previous RC scheme,[42] the charge on each M1 atom is redistributed to the midpoints of M1–M2 bonds. However, the total charge of the QM/MM entire system is not conserved when the sum of MM charges of the MM region is not zero or an integer. In the balanced RC scheme, introduced here, we first adjust the charge on the M1 atom so that

$$q_0 + \sum_i q_i + q^{\mathrm{QM}} = q^{\mathrm{total}} \tag{2.4}$$

where q_0 is the modified M1 charge, $\{q_i\}$ are the MM point charges of other MM atoms (except M1), q^{QM} is the charge of the QM region (that is, of the capped QM subsystem),

q^{total} is the charge of the original entire system. This charge balancing step conserves the charge of the QM/MM entire system.

Then the balanced RC scheme redistributes the charge q_0 evenly to the midpoints of all M1−M2 bonds, with each bond midpoint obtaining a charge $q^{\text{RC}} = \frac{q_0}{n}$, where *n* is the number of M1−M2 bonds. For the balanced redistributed charge and dipole (balanced RCD) method, we double the redistributed charges and adjust the charges q_{M2}^{RCD} on M2 atoms to conserve the total charge of the QM/MM entire system, as shown in eqs 2.5 and 2.6:

$$q^{\text{RCD}} = 2q^{\text{RC}} = \frac{2q_0}{n} \tag{2.5}$$

$$q_{M2}^{\text{RCD}} = q_{M2} - q^{\text{RC}} \tag{2.6}$$

These two schemes are illustrated in Figure 2.1.

In this study, we compare balanced RC and balanced RCD to other methods that differ in how the redistributed charges are handled, e.g., to what location are they redistributed. These methods include: balanced straight electrostatic embedding (SEE), balanced RC2, Amber-1,[43] balanced RC3, Amber-2,[43] and balanced Shift.[44] Amber-1 and Amber-2 are the options called *adjust_q* = 1 and *adjust_q* = 2 in *AMBER 10*. The distinction between these methods is in the position of the redistributed charges and whether the dipoles of M1−M2 bonds are corrected. In balanced SEE, the charge on the M1 atom is set to q_0, and it is not moved. In balanced RC2, we distribute q_0 evenly to all M2 atoms. In balanced RC3, we distribute q_0 evenly to all M2 and M3 atoms. In Amber-1, we move q_0 to the nearest M2 atom. In Amber-2, we distribute q_0 evenly to all MM

atoms, except the M1 atom. (Note that Amber-2 is the default option in revision 10 of AMBER,[46] whereas Amber-1 can be selected in AMBER 10 by specifying *adjust_q* = 1) In balanced Shift, the redistributed charges are placed at M2 atoms, and dipoles are added around M2 atoms to compensate the movement of the charges. A summary of these charge schemes is shown in Table 2.1.

We call the methods in Table 2.1 balanced methods because they all conserve the total charge of the QM/MM entire system. Five unbalanced methods, in which the total charge of the QM/MM entire system is not necessarily conserved, are also tested, including SEE, Z1, Z2, Z3, and RC.[42] SEE is straight electronic embedding that makes no change of the charges of MM boundary atoms, Z1 denotes that the charge of the M1 atom is zeroed (this can be chosen by specifying *adjust_q* = 0 in AMBER 10, and it is the default method in CHARMM[47]), Z2 denotes that the charges of M1 and M2 atoms are zeroed (Z2 is the default scheme in both *Gaussion03*[48] and *Gaussian09*[49]), and Z3 denotes that all the charges of all M1, M2, and M3 atoms are zeroed. RC denotes that the charge on the M1 atom is redistributed to the midpoints of M1–M2 bonds without the balancing step. Balanced methods and unbalanced methods are compared to test the importance of conserving the charge of the QM/MM entire system. To make a comparison, we also carry out calculations on the capped primary system (CPS), in which the whole MM region is substituted by the link atom.

2.2.2 Link atom

Another issue in the boundary treatment is the choice of the link atom. A hydrogen atom can be used as the link atom when a C–C bond is cut. However, a Q1–H bond may be a poor model for the cut bond when the M1 atom is electronegative, such as in a Si–O

or C–O bond. Therefore, we use a tuned capping atom as the link atom to mimic a cut polar bond and to reproduce the electronic structure of the QM subsystem. Redondo et al.[29] used a tuned hydrogen atom to replace a silicon atom. Koga et al.[30] added a shift operator on the hydrogen atom to reproduce the effect of the substitution. Zhang et al.[33] and Nasluzov et al.[31] used tuned fluorine atoms and derived pseudopotentials for carbon and oxygen boundary atoms. Here, we provide a more general rule to tune a capping atom for boundary atoms. The capping atom is always a tuned F atom. We first replace the $1s^2$ core by a conventional pseudopotential U, and then a tuning pseudopotential $U_0'(r)$ is added to U. The conventional pseudopotential used here is the CRENBL effective core potential (ECP) for a fluorine atom developed by Pacios and Christiansen.[50] The form of this potential is

$$U = U_0(r) + \sum_{m=-1}^{1} [U_1(r) - U_0(r)] |1m\rangle\langle m1| \tag{2.7}$$

where

$$U_l(r) = r^{-2} \sum_j C_{lj} r^{n_{lj}} e^{-\alpha_{lj} r^2} \tag{2.8}$$

r is the distance of an electron from the capping nucleus, and $|lm\rangle$ is a spherical harmonic. The parameters for this pseudopotential are listed in Table 2.2. The form of $U_0'(r)$ is

$$U_0'(r) = C \exp[-(r/r_0)^2] \tag{2.9}$$

where C and r_0 are parameters. The basis set used for the tuned F atom is the same as for a conventional F atom. For example, if the QM subsystem is treated by the 6-31G* basis set, then the tuned F atom has the 6-31G* basis set of a conventional F atom. To find an

appropriate pseudopotential, we set r_0 equal to one bohr, and tune the parameter C of the pseudopotential.

The next key decision is how to choose C. In order to reproduce the electronic structure of the QM subsystem, we require that the total charge of the uncapped portion of the QM subsystem in a QM/MM calculation is equal to the total charge of the same subsystem in a QM calculation of the original entire system or, in practice, of a system that mimics the original entire system better than the capped QM subsystem does (see below for more details of this large system). Mulliken charges were used as the indicator. Because Mulliken charges become unphysical when large basis sets are used, we used small basis sets without diffuse functions for this tuning step, in particular, 6-31G* when M1 is from the second period (Li through F) and STO-3G otherwise. Since the STO-3G basis set is defined for the entire periodic table, the tuning step is well defined for the entire periodic table.

We can perform the tuning process on either the reactant or the product. For the validation suite, the reactant is a neutral molecule, and the product is a deprotonated anion. In this study, we used the protonated neutral reactant to tune the F atom. All the parameters C of the pseudopotentials are derived in the presence of MM background charges. Because the MM charges are redistributed differently in the various boundary charge schemes explained in section 2.2.1, the derived pseudopotentials are not the same for different charge schemes.

To enable the method to be applied to large systems for which it is difficult to perform a QM calculation on the original entire system, the tuning is performed on a model system created from the original entire system. This model system is called the

tuning system or the entire system model (ESM). It consists of all QM atoms and all M1, M2, and M3 atoms, and all free valences on M3 atoms are capped by untuned H atoms. The scheme is illustrated in Figure 2.2.

The completely defined tuning process employed in the present study is as follows:

1. Choose a geometry and charge state for the tuning system and create the entire system model (ESM) by capping all M3 atoms with untuned hydrogens.

2. Do a full QM calculation on ESM and carry out Mulliken population analysis. For the basis set, use 6-31G* if M1 is from the second period (e.g., C, N, O), and use STO-3G otherwise (that is, if M1 belongs to the 3rd or higher period, for example, Si). This yields the total charge on the primary system (PS); call this $q_{PS}^{\mathrm{ESM,MPA}}$, where MPA denotes Mulliken population analysis. It also yields $q_{SS}^{\mathrm{ESM,MPA}}$, where SS is the secondary subsystem, equal to all the atoms in ESM except the PS atoms. By construction, $q_{PS}^{\mathrm{ESM,MPA}} + q_{SS}^{\mathrm{ESM,MPA}} = q_{ESM}$ where q_{ESM} is an integer.

3. Select an MM charge scheme. For the present calculations, the MM charge scheme is CM4M charges from a calculation on the ESM. The basis set used for the calculation of CM4M charges could in principle be the same as chosen for step 2, but in fact, we do not have CM4M charge schemes for STO-3G; therefore the MM charges are always CM4M charges determined with M06-2X/6-31G* calculations on the ESM.

4. Define TSS as the truncated secondary system of the original entire system model, which includes all atoms in the secondary subsystem of the ESM except the M1 atom. For the chosen MM charge scheme of step 3, calculate $q_{TSS}^{\mathrm{ESM,MM}}$, which equals the sum of the MM charges from step 3 (that is, the sum of CM4M charges) on all TSS atoms of ESM.

5. Cap the PS with an F* atom to create the capped primary system (CPS), where F* denotes a tuned F atom. Always set r_0 equal to 1 bohr in the pseudopotential. The other parameter (C) of the pseudopotential will be determined in step 7.

6. Select a charge modification scheme, for example, balanced RC or balanced RCD. For the balanced charge schemes, we set q_0 to make $q_{CPS} + q_0 + q_{TSS}^{\mathrm{ESM,MM}}$ equal to q_{ESM}. In the usual case where $q_{CPS} = q_{ESM}$, then this yields: $q_0 = -q_{TSS}^{\mathrm{ESM,MM}}$.

7. Now, for a given MM charge scheme, and given charge modification scheme, carry out a series of fixed-geometry CPS** calculations with various values of C. Note that CPS** here denotes the capped primary system in the modified charge environment of the secondary system of the ESM. Adjust C until q_{F*}^{MPA} equals $q_{SS}^{\mathrm{ESM,MPA}}$, which was determined in step 2. Now the pseudopotential is known, so F* is properly tuned.

After the tuning step, the tuned F link atom can be used with the selected charge scheme on the QM/MM entire system to do QM/MM calculations on the proton affinities.

2.3 Details of validation calculations

We have implemented the proposed charge schemes and link atom treatments in the QMMM package,[51] which is based on the *Gaussian03*[48] and TINKER[52] programs. Either density functional theory (DFT), or wave function theory (WFT) can be used for the QM calculations. In the study, the M06-2X density functional[53,54] was used for all the QM calculations. Proton affinities were used as the criterion to evaluate the methods, as they are sensitive to the boundary treatment.[47] The 6-31G** basis set was used for

aluminosilicate clusters, and the 6-31G* basis set was used for organic molecules.

The geometry of all the molecules was fixed at the protonated geometry, so in the eqs 2.1 and 2.2, the QM/MM valence term $E_{val}^{\mathrm{QM/MM}}$, the QM/MM van der Waals term $E_{vdW}^{\mathrm{QM/MM}}$ and the MM term E^{MM} are the same for the deprotonated and protonated forms, and they cancel out in the QM/MM calculations of proton affinities. The final expression for the proton affinity is

$$\begin{aligned} E(\text{proton affinity}) &= E(\text{deprotonated}) - E(\text{protonated}) \\ &= [E^{\mathrm{QM}}(\text{deprotonated}) + E_{el}^{\mathrm{QM/MM}}(\text{deprotonated})] \\ &\quad - [E^{\mathrm{QM}}(\text{protonated}) + E_{el}^{\mathrm{QM/MM}}(\text{protonated})] \end{aligned} \tag{2.10}$$

In this study, CM4M charges[55] were used for MM atoms. The protonated forms of the molecules in the test suite are illustrated in Figure 2.3. In each case, the QM region is on the left, and the MM region is on the right. The selected molecules in the test suite contain different kinds of Q1–M1 bonds at the QM/MM boundary, in particular, C–C, N–C, O–C, S–C, C–N, O–N, C–O, Al–O, Si–O, C–Si, O–Si, C–S, and S–S. Both polar and nonpolar bonds are included in this test suite.

For the position of the link atom, the scaled bond distance method[56,57] was used, that is, the link atom is placed along the Q1–M1 bond, and the ratio of the Q1–link atom distance to the Q1–M1 distance is set to be the ratio of the standard bond length of the Q1–link atom bond to the standard bond length of the Q1–M1 bond. The standard bond lengths used in this study are listed in Table 2.3. For the tuned F link atom, it is placed at the same position as that of an ordinary F atom.

2.4 Results

2.4.1 An example for balancing the charge and tuning the link atom

We use molecule **w** as an example to demonstrate the proposed algorithm. The O–Si bond is cut. In the tuning process, we create the entire system model (ESM) from the original QM system. The original entire system and the ESM are shown in Figure 2.4. A full QM calculation is performed on the ESM to get the total Mulliken charge on the QM region $q_{\mathrm{PS}}^{\mathrm{ESM,MPA}}$ and on the MM region $q_{\mathrm{SS}}^{\mathrm{ESM,MPA}}$. As the M1 atom is silicon, STO-3G basis set was used. CM4M charges were calculated for all the MM atoms in ESM; the M1 atom has a charge of 0.515e, and the sum $q_{\mathrm{TSS}}^{\mathrm{ESM,MM}}$ of the MM charges in the truncated secondary system (TSS) is –0.377e. Therefore, in the balanced schemes, the redistributed charge $q_0 = -\, q_{\mathrm{TSS}}^{\mathrm{ESM,MM}} = 0.377e$. Then the QM system is capped with a tuned F atom, and the capped QM subsystem is embedded in the redistributed MM charges using a boundary charge scheme. The parameter C of the pseudopotential is adjusted to make the Mulliken charge $q_{\mathrm{F}*}^{\mathrm{MPA}}$ of the tuned F atom equal to $q_{\mathrm{SS}}^{\mathrm{ESM,MPA}}$. For example, in the balanced RC scheme, the parameter of the pseudopotential is 0.80, as shown in Table 2.6.

The tuned F* link atom is used to cap the QM system in the QM/MM entire system. The same charge scheme is used for the tuning and the calculations of proton affinities.

The results with the tuned F atom are compared to those with untuned H and F link atoms.

2.4.2 H and F link atoms

Tables 2.4 and 2.5 show the proton affinities of the twenty-five molecules by full QM calculations and the signed error by QM/MM calculations using untuned H and F atoms as link atoms.

2.4.2.1 Balanced and unbalanced charge schemes

The balanced methods (balanced SEE, balanced RC, Amber-1, balanced RC2, balanced RC3, Amber-2, balanced RCD, balanced Shift) give much smaller errors in proton affinities than the unbalanced ones (SEE, Z1, Z2, Z3, RC). Both the H link atom and the F link atom schemes have the same trends. The mean unsigned errors (MUEs) given by all the balanced methods are 4–7 kcal/mol, while the MUEs given by all the unbalanced methods are 15–24 kcal/mol. This is because in the unbalanced methods, a net partial change is created near the QM region and the interactions between QM and MM regions become unphysical.[4,33,43] The CPS method, in which any polarization of the QM region by the MM region is excluded, doesn't change the total charge of the QM/MM entire system, and gives a smaller MUE than the unbalanced methods. Therefore, the conservation of the total charge of the QM/MM entire system is one of the key factors for the calculations of proton affinities.

2.4.2.2 Different link atoms and charge schemes

The comparison of the results using the H link atom (Table 2.4) and the F link atom (Table 2.5) shows that the proton affinities are sensitive to the link atom. In most cases, the H link atom gives larger proton affinities than the full QM calculations, while the F link atom gives smaller proton affinities. As the link atom is directly connected to the QM region, it can greatly change the electronic structure of the QM region. The Q1–H and Q1–F bonds cannot adequately reproduce the properties of the original Q1–M1 bond. Electronegativity can be used as a qualitative criterion to decide which atom is better to be used as the link atom. For example, when the M1 atom is oxygen or nitrogen, the F link atom gives better results than the H link atom.

Moreover, compared with the charges on the other MM atoms, the charge on the link atom is closest to the QM region and greatly affects the electrostatic potential on the active site. König et al.[47] have compared different charge schemes to treat the boundary and found that all the balanced methods give, on average, comparable errors in proton affinities and deprotonation energies. Here, we also found that in the proton affinity calculations, all balanced charge schemes give similar mean unsigned errors (MUEs) of twenty-five molecules in the test suite. As the link atom affects the charge distribution near the QM/MM boundary, it is possible that the errors brought by the different charge schemes and by the link atom are of the same order of magnitude. In the next section, we will first tune the link atom to make the total charge of the QM subsystem right , and then compare different charge schemes.

The position of the active site for each molecule is also listed in the Tables 2.4 and 2.5. We found that even when the active site is far from the boundary, the error is still

large if the boundary is not well treated. For example, in molecule **v**, the active site is at the Q5 atom, but the errors range from –5.9 to 10.4 kcal/mol using different link atoms and different balanced charge schemes. Therefore, increasing the distance of the active site from the boundary cannot completely remove the error due to the link atom. When a polar bond is cut, a tuned link atom and balanced charge scheme should be used.

2.4.3 Tuned F link atom

We used the protocol presented in section 2.2.2 to tune the F link atoms. The final parameters *C* of the pseudopotentials for the tuned F atoms are shown in Table 2.6. These parameters reflect the differences among various kinds of bonds. The same type of bond shows similar parameters even in different molecules. For example, in the balanced RC scheme, the parameter for the pseudopotential is 0.65 ~ 1.45 for a carbon boundary atom, –0.15 ~ 0.60 for a nitrogen boundary atom, –0.45 ~ –0.15 for an oxygen boundary atom, 0.60 ~ 0.80 for a silicon boundary atom, and 0.30 ~ 0.40 for a sulfur boundary atom. These results are consistent with the electronegativities of the atoms. Also, we found that when the M1 atom is less electronegative than the Q1 atom, the pseudopotential for the tuned F atom is larger. For example, the pseudopotential needed for an O–N bond (0.60) is much larger than that needed for a C–N bond (–0.15).

The tuned F atom was used as the link atom to calculate the proton affinities, and the results are shown in Table 2.7. By examination of the mean unsigned error (MUE) of the twenty-five molecules, one finds that the tuned F link atom gives smaller errors than the H link atom (Table 2.5) or the F link atom (Table 2.6) in *all* the charge schemes. This indicates that the accurate treatment of the boundary is very important, especially after

the total charge is conserved. The tuned F link atom makes the total charge of the QM region right, and it avoids the artifacts that can be introduced by the use of a link atom.

The balanced RC scheme gives the best results, and the MUE is only 1.3 kcal/mol. The good performance of the balanced RC scheme can be explained as follows. In order to correctly handle the electrostatic interactions between the QM region and the MM region in the QM/MM calculations, it is important to have an accurate charge distribution of the MM region. When we move the boundary charges to avoid overpolarization, the charge distribution of the MM region can be greatly changed if the redistributed charges are moved far from the boundary. In the balanced RC method, the redistributed charges are moved to the midpoints of the M1–M2 bonds, and, compared to other charge schemes, they introduce smaller changes to the charge distributions in the MM regions. In the balanced RC2, RC3, and Amber 2008-2 schemes, the redistributed charges are placed farther and farther from the boundary region, and the MUE of the proton affinities increases. When we use balanced RC3, in which the charges are redistributed to M2 and M3 atoms, the error is approximately equal to that in the capped primary system (CPS). However, when the bond is non-polar, such as a C–C bond, the redistributed charge is relatively small and different charge schemes give similar errors.

In all three tables of mean unsigned errors (Tables 2.4, 2.5, and 2.7), the balanced RC scheme is better than the balanced RCD scheme, whereas previously[42] the RCD scheme performed slightly better. Since the testing is more thorough in the present paper, we now believe that that the simpler RC scheme is to be preferred to the RCD one.

From the above results, we conclude that both a good charge scheme and an appropriate treatment of the link atom are needed to accurately treat the boundary. The balanced RC method with a tuned F atom gives the best results among all the methods.

However, there are still some problems. For example, we found that the error for the C–S bond is still quite large, and there is no obvious reason for this error.

2.4.4 Tuning the F link atom based on other methods

Having established that we can obtain better results with a tuned fluorine atom, we should note that in future work one could also consider other ways to do the tuning. For example, the tuned parameter could be tuned to make the proton affinity or a particular reaction energy come out right at representative geometries of the reactant or product. There would be three advantages of this approach: (i) the proton affinity, unlike the partial charges, is a physical observable, that can be calculated straightforwardly; (ii) the proton affinity depends on both the initial and final states, so one does not have to make a decision whether to tune the pseudopotential in the reactant state or the product state; (iii) there is additional flexibility in that one could tune the calculated proton affinity or reaction energy either (a) to a calculation in which a portion of the MM system (e.g., the M1, M2, and M3 atoms or the functional groups that contain them) is treated quantum mechanically, but at the same level as is to be used for the QM portion of the QM/MM calculations, or (b) to experiments or high–level calculations.

2.5 Concluding remarks

In QM/MM studies it is often inevitable to cut covalent bonds between the QM and MM parts. The present study addresses QM/MM boundary treatments in a systematic

manner. All the ingredients considered are well known in the literature from our own work and that of others (relevant references are cited above), in particular, the use of link atoms, the balancing and redistribution of charges close to the QM/MM boundary, and the tuning of the properties of the link atoms by suitable calibration (e.g., via a pseudopotential). In the present article we present a new method of tuning, and we combine it with the other ideas just mentioned in a new way to yield a method called the balanced redistributed charge (BRC) scheme with tuned fluorine link atom. If an acronym is needed to save space, this could be labeled the TBRC (tuned and balanced redistributed charge) method.

We apply the new method to calculate the proton affinities of 25 diverse compounds (at fixed geometries), and we compare its performance to that of other boundary charge and link atom schemes for treating the QM/MM boundary with regard to their ability to reproduce the quantum mechanical reference values. The balanced redistributed charge scheme with tuned fluorine link atoms outperforms the other treatments for the chosen validation suite of proton affinities, as shown in Tables 2.4, 2.5, and 2.7. Some of the methods listed in Table 2.4 are typical methods currently in use. For example, as mentioned in section 2.2.1, the Z1 scheme is the default in CHARMM[47], the Z2 scheme is the default in both the *Gaussian03*[48] and *Gaussian09*[49] packages, and the Amber 2008-2 method is the default in AMBER-version 10.[46]

One important finding is that the errors are generally larger for treatments without charge balancing, whereas the choice of the actual charge redistribution scheme is less crucial, but not insignificant. From Tables 2.4 and 2.5, we see that the mean unsigned error (MUE) in the unbalanced schemes ranges from 15 to 24 kcal/mol, much worse than

even the CPS method, which has no MM region (just a capped quantum subsystem) and which gives an MUE of 5 kcal/mol. The importance of balancing, which was previously emphasized by others,[33,43-45] has a dramatic effect, reducing the MUE to 4–7 kcal/mol.

A key physical element of the new scheme is that we tune the link atom to try to make the charge on the primary subsystem (the atoms of the capped primary system, but excluding the cap) be the same as it would be in a quantum mechanical calculation on the entire system. In the past there has been much more emphasis on redistributing the MM charge near the boundary than on the charge on the primary subsystem. If, however, the charge on the uncapped portion of the capped quantum mechanical system is inaccurate, then no treatment of the boundary can restore the correct physics. This motivation for tuning is substantiated by the finding that the use of a tuned fluorine link atom further reduces the mean unsigned error in all eight balanced charge schemes, leading to mean unsigned errors of 1–4 kcal/mol. In six of these eight cases, the MUE is smaller that that in CPS with a tuned F atom. All these eight schemes give smaller MUEs than CPS capped with a hydrogen atom. The MUE of 1.3 kcal/mol for the combination of the balanced RC scheme for boundary charges and the tuned fluorine atom is particularly encouraging in light of the difficulty of the test set. In fact, the average proton affinity in all 25 cases of the test set is 373 kcal/mol, and the range of quantum mechanical reference values is 67.5 kcal/mol (from 339.1 to 406.6 kcal/mol). The mean unsigned error of 1.3 kcal/mol for the recommended new methods is only 1.9% of the range.

We conclude that two elements are very important in the energy calculations: balancing the charges of the MM region and tuning the link atom. A general rule (defined for all single bonds in the whole periodic table) is provided to tune the link atom when

different types of single bonds are cut at the boundary. We can calculate accurate proton affinities after correctly handling the QM/MM boundary.

In the present work, motivated by interest in a variety of catalytic and redox systems, we intentionally chose a very difficult property, proton affinities, we considered QM–MM boundaries both close and far from the site of protonation, and we created a test set that is much more diverse than previous test sets used for QM/MM methods, in particular in that it includes boundaries that cut very polar bonds, including some in which neither atom is a carbon. Some of the methods found to be inadequate for this demanding test set will perform better for properties that are less sensitive than proton affinities to electrostatic potentials in the quantum subsystem or for cases where only a carbon–carbon bond is broken. However, simply enlarging the quantum system enough to move the QM–MM boundary far from the site of reaction, when affordable, is not sufficient to guarantee good accuracy. Furthermore, in simulations of complex systems, it is often desirable to use a method that has been validated for even the most challenging problems.

In future work, we will examine the problem of geometry optimization using the new QM/MM scheme. In the present article, the valence and van der Waals terms that involve interactions between the QM subsystem and MM subsystem cancel out, but we will need a protocol for defining them when we begin to optimize geometries or consider dynamics.

2.6 References for Chapter 2

1. Truhlar, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 16824.
2. Gao, J. *Rev. Comp. Chem.* **1996**, *7*, 119.
3. Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

4. Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; Neuman Institute for Computing; Jülich, Germany, 2000; Vol. 3, p 285
5. Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc.* **2001**, *106*, 105.
6. Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Röhrig, U.; Spiegel, K.; Sulpizi, M.; VandeVondele, J.; et al. *Chimia* **2002**, *56*, 13.
7. Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
8. Shurki, A.; Warshel, A. In *Protein Simulations*; Vol. 66; Academic Press Inc: San Diego, USA, 2003; p. 249.
9. Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1.
10. Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
11. Hammes-Schiffer, S. *Curr. Opin. Struct. Biol.* **2004**, *14*, 192.
12. Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
13. Moret, M.-E.; Tapavicza, E.; Guidoni, L.; Röhrig, U.; Sulpizi, M.; Tavernelli, I.; Rothlisberger, U. *Chimia* **2005**, *59*, 493.
14. Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393.
15. Jorgensen, W. L.; Tirado-Rives, J. *J. Comp. Chem.* **2005**, *26*, 1689.
16. Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comp. Chem.* **2005**, *26*, 1701.
17. Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; et al. *J. Phys. Chem. B* **2006**, *110*, 6458.
18. Zhang, Y. *Theor. Chem. Acc.* **2006**, *116*, 43.
19. Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182.
20. Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
21. Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573.
22. Sousa, S. F.; Ramos, M. J. In *Computational Proteomics 2008*; Ramos, M. J., Ed.; Transworld Research Network: Kerala, India, 2008; p. 101.
23. Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed. Engl.* **2009**, *48*, 1198.
24. Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1986**, *7*, 718.
25. Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
26. Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comp. Chem.* **1994**, *15*, 269.
27. Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
28. Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.
29. Redondo, A.; Goddard, W. A.; Swarts, C. A.; McGill, T. C. *J. Vac. Sci. Technol.* **1981**, *19*, 498.
30. Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1990**, *172*, 243.
31. Nasluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228.
32. Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
33. Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.

34. Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
35. Parks, J. M.; Hu, H.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 154106.
36. Alary, F.; Poteau, R.; Heully, J. L.; Barthelat, J. C.; Daudey, J. P. *Theor. Chem. Acc.* **2000**, *104*, 174.
37. DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
38. Moon, S.; Christiansen, P. A.; DiLabio, G. A. *J. Chem. Phys.* **2004**, *120*, 9080.
39. DiLabio, G. A.; Wolkow, R. A.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 044708.
40. von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.
41. Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
42. Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.
43. Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comp. Chem.* **2008**, *29*, 1019.
44. Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79.
45. Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
46. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; and Kollman, P. A.; *AMBER 10*; University of California: San Francisco, CA, 2008.
47. König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.
48. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03*, revision D. 01; Gaussian, Inc.: Wallingford, CT, 2004.
49. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.;

Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.

50. Pacios, L. F.; Christiansen, P. A. *J. Chem. Phys.* **1985**, *82*, 2664.
51. Lin, H.; Zhang, Y.; Truhlar, D. G. QMMM; Version 1.3.5; University of Minnesota: Minneapolis, 2007.
52. Ponder, J. W. Tinker; Version 4.2; Washington University: St. Louis, MO, 2004.
53. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
54. Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
55. Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046.
56. Maseras, F.; Morokuma, K. *J. Comp. Chem.* **1995**, *16*, 1170.
57. Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. Theochem* **1999**, *461*, 1.

Table 2.1 Charge schemes

	Position of the redistributed charges	Correction of bond dipole	Reference
balanced SEE	M1 atom	no	present
balanced RC	midpoints of M1–M2 bonds	no	present
balanced RC2	M2 atoms	no	present
Amber-1	Nearest M2 atom	no	Walker et al.[43]
balanced RC3	M2, M3 atoms	no	present
Amber-2	all MM atoms (except M1 atom)	no	Walker et al.[43]
balanced RCD	midpoints of M1–M2 atoms	yes	present
balanced Shift	M2 atoms	yes	Sherwood et al.[44]

Table 2.2 The CRENBL effective core potential[a]

	n_{lj}	α_{lj}	C_{lj}
U_0	2	2.8835	12.685306
	2	3.1077	–19.302589
	1	5.6122	1.002179
	0	2.8146	2.245349
U_1	2	44.5166	–6.723024
	2	12.9487	–0.929649
	1	132.4967	–1.526734

[a]Ref. 50

Table 2.3 Standard bond lengths (Å)

bond	distance	bond	distance	bond	distance	bond	distance
C–H	1.09	C–F	1.33	C–C	1.53	N–O	1.47
N–H	1.01	N–F	1.41	N–C	1.45	Al–O	1.72
O–H	0.95	O–F	1.41	O–C	1.42	Si–O	1.61
Al–H	1.55	Al–F	1.67	S–C	1.81	S–S	2.04
Si–H	1.45	Si–F	1.56	Si–C	1.86		
S–H	1.34	S–F	1.65				

Table 2.4 Full-QM proton affinities (PA, in kcal/mol), QM/MM signed errors (in kcal/mol), and mean unsigned errors (MUE) (in kcal/mol) averaged over 25 cases using the H link atoms

case	bond	PA	site	CPS	bal. SEE	bal. RC	bal. RC2	Amber-1	bal. RC3	Amber-2	bal. RCD	bal. Shift	SEE	Z1	Z2	Z3	RC
a	C-C	362.3	Q3	6.8	1.2	2.5	3.9	3.5	3.9	3.9	1.0	1.3	-2.4	13.4	6.8	6.8	-0.7
b	C-C	406.6	Q3	-0.6	3.9	5.1	6.3	5.5	6.9	7.3	3.9	4.2	4.8	13.1	-3.6	-40.7	5.8
c	C-C	402.6	Q2	9.0	12.0	13.0	14.0	13.7	14.5	14.5	11.9	12.1	12.0	19.7	-19.2	9.0	13.0
d	N-C	376.7	Q4	3.4	6.6	4.8	2.9	3.8	1.7	1.2	6.6	6.2	-14.7	-9.7	7.2	-15.1	-14.1
e	O-C	398.6	Q4	4.0	5.4	4.1	2.9	3.4	2.0	1.4	5.3	5.0	-7.8	-5.7	35.5	-26.7	-7.5
f	S-C	394.7	Q4	-1.1	2.2	0.6	-1.2	-0.5	-2.6	-3.2	2.3	1.9	-6.8	-15.0	2.0	-16.4	-7.6
g	C-N	355.2	Q3	13.0	17.0	12.5	7.4	11.3	2.5	-1.4	17.5	16.2	39.9	-26.9	47.0	-3.6	32.9
h	O-N	400.0	Q4	3.0	13.8	9.6	5.4	5.2	3.8	3.8	13.5	12.4	6.8	-19.1	-23.0	3.0	3.4
I	C-O	394.9	Q3	11.9	12.1	11.3	10.3	10.3	9.7	9.3	12.4	12.1	38.8	5.9	40.2	-5.9	34.4
j	C-O	401.0	Q3	5.5	13.9	11.1	7.7	7.7	6.1	5.1	14.4	13.5	37.5	-6.8	12.0	-30.5	31.3
k	C-O	366.7	Q3	5.8	10.8	7.5	4.2	4.2	2.5	1.6	10.8	10.0	27.0	-11.8	9.2	-35.4	21.3
l	C-O	398.3	Q4	7.1	7.6	7.1	6.5	6.5	6.2	5.9	7.7	7.6	28.1	3.5	30.6	-7.8	25.1
m	Al-O	340.7	Q2	5.1	6.0	3.8	1.1	1.1	-0.2	-0.3	6.5	5.7	33.6	-10.8	19.9	-18.2	28.5
n	Al-O	339.1	Q2	6.8	6.0	3.8	1.0	1.0	-0.4	-0.3	6.6	5.7	33.6	-11.1	26.1	-42.0	28.4
o	Al-O	348.0	Q2	-2.2	5.5	3.7	1.5	1.5	0.4	-0.2	5.9	5.1	35.1	-8.0	35.1	-61.5	29.9
p	Si-O	349.0	Q2	-4.9	4.0	0.7	-2.9	-2.9	-4.5	-6.3	4.3	2.8	27.2	-15.9	27.9	-35.5	20.1
q	Si-O	353.2	Q4	-4.0	0.7	-0.4	-1.8	-1.8	-2.5	-3.8	1.0	0.7	18.3	-11.0	22.8	-22.2	15.6
r	Si-O	348.1	Q2	-4.1	4.0	0.6	-3.2	-3.2	-5.0	-6.8	4.4	2.8	23.9	-17.3	30.7	-60.6	17.4
s	C-Si	397.0	Q3	8.6	-2.9	2.3	7.2	6.2	7.2	7.2	-3.0	-1.3	-20.4	33.4	8.6	8.6	-12.6
t	O-Si	342.7	Q3	5.8	3.9	4.9	5.8	6.0	5.8	5.8	4.0	4.3	-8.1	13.2	5.8	5.8	-5.7
u	O-Si	348.0	Q3	0.5	1.7	6.2	9.8	10.8	11.7	11.7	2.4	3.5	-9.1	38.7	-59.5	0.5	-3.1
v	O-Si	353.2	Q5	1.6	2.2	4.6	6.6	7.9	8.2	10.4	2.4	2.9	-8.1	29.2	-16.8	24.5	-4.7
w	O-Si	354.8	Q5	0.0	2.7	5.3	7.5	9.0	9.3	14.2	3.0	3.4	-5.0	32.5	-13.0	38.6	-1.6
x	C-S	395.8	Q3	10.5	16.1	14.4	12.6	12.6	12.1	12.1	16.2	15.5	25.5	5.4	-7.5	10.5	22.3
y	S-S	390.3	Q4	2.8	6.1	5.1	4.0	4.0	3.7	3.7	6.1	5.8	5.3	-2.8	-14.8	2.8	4.4
MUE				5.1	6.7	5.8	5.5	5.7	5.3	5.7	6.9	6.5	19.2	15.2	21.0	21.3	15.7

Table 2.5 Full-QM proton affinities (PA, in kcal/mol), QM/MM signed errors (in kcal/mol), and mean unsigned errors (MUE) (kcal/mol) of 25 cases by QM/MM calculations using F link atoms

case	bond	PA	site	CPS	bal. SEE	bal. RC	bal. RC2	Amber-1	bal. RC3	Amber-2	bal. RCD	bal. Shift	SEE	Z1	Z2	Z3	RC
a	C-C	362.3	Q3	-2.6	-7.8	-7.0	-5.6	-6.0	-5.6	-5.6	-8.4	-8.1	-11.1	3.8	-2.6	-2.6	-10.1
b	C-C	406.6	Q3	-10.4	-5.4	-4.6	-3.4	-4.2	-2.9	-2.5	-5.7	-5.4	-4.6	3.2	-12.4	-49.7	-3.9
c	C-C	402.6	Q2	-8.5	-5.3	-4.8	-3.7	-4.0	-3.3	-3.3	-5.8	-5.6	-5.3	1.8	-36.3	-8.5	-4.8
d	N-C	376.7	Q4	-11.5	-8.2	-9.6	-11.7	-10.8	-13.0	-13.5	-7.4	-7.7	-29.6	-24.7	-7.1	-30.3	-29.2
e	O-C	398.6	Q4	-10.4	-9.7	-10.7	-12.3	-11.7	-13.3	-14.0	-9.2	-9.4	-23.7	-21.5	22.6	-42.3	-23.5
f	S-C	394.7	Q4	-7.4	-3.7	-5.2	-7.3	-6.5	-8.8	-9.4	-3.1	-3.7	-12.9	-21.2	-4.0	-22.7	-13.6
g	C-N	355.2	Q3	4.0	5.4	3.4	-1.5	2.3	-6.2	-9.9	8.4	7.1	26.8	-34.7	37.4	-12.5	23.6
h	O-N	400.0	Q4	-11.0	2.2	-1.8	-7.7	-7.9	-9.8	-9.8	3.9	1.7	-5.5	-33.8	-37.9	-11.0	-8.7
I	C-O	394.9	Q3	3.1	3.3	2.9	1.9	1.9	1.3	1.0	3.8	3.6	27.6	-2.3	30.5	-14.3	25.1
j	C-O	401.0	Q3	-2.8	4.6	2.8	-0.6	-0.6	-2.1	-3.0	6.2	5.0	26.6	-14.5	3.5	-37.8	22.8
k	C-O	366.7	Q3	-0.8	3.5	1.4	-2.3	-2.3	-4.2	-5.1	5.2	4.0	19.2	-18.4	2.7	-41.9	15.7
l	C-O	398.3	Q4	1.5	1.8	1.5	0.9	0.9	0.6	0.3	2.1	1.9	21.0	-2.1	24.4	-13.3	19.3
m	Al-O	340.7	Q2	2.1	2.7	1.0	-1.8	-1.8	-3.1	-3.2	3.8	2.7	30.3	-13.5	16.6	-20.8	25.6
n	Al-O	339.1	Q2	3.8	2.7	0.9	-1.9	-1.9	-3.3	-3.2	3.8	2.7	30.3	-13.9	22.7	-44.5	25.5
o	Al-O	348.0	Q2	-5.2	2.0	0.6	-1.6	-1.6	-2.7	-3.2	2.9	2.0	31.4	-11.0	31.5	-63.9	26.8
p	Si-O	349.0	Q2	-8.5	-0.2	-2.8	-6.6	-6.6	-8.2	-10.1	1.0	-0.7	22.5	-19.5	24.0	-39.0	17.0
q	Si-O	353.2	Q4	-3.3	1.3	0.4	-1.2	-1.2	-2.0	-3.4	2.0	1.6	19.3	-10.5	23.7	-21.5	16.9
r	Si-O	348.1	Q2	-6.9	0.6	-2.1	-6.2	-6.2	-8.1	-9.8	2.0	0.1	20.3	-20.2	27.6	-62.9	15.0
s	C-Si	397.0	Q3	-1.1	-11.9	-7.2	-2.5	-3.6	-2.5	-2.5	-11.8	-10.0	-27.8	23.1	-1.1	-1.1	-21.2
t	O-Si	342.7	Q3	-3.8	-6.7	-5.1	-3.8	-3.6	-3.8	-3.8	-6.5	-5.9	-20.1	3.9	-3.8	-3.8	-16.7
u	O-Si	348.0	Q3	-9.1	-11.0	-4.4	1.1	2.2	3.2	3.2	-9.8	-7.1	-22.5	31.2	-70.8	-9.1	-14.4
v	O-Si	353.2	Q5	-4.4	-5.9	-2.3	0.8	2.0	2.6	5.1	-5.4	-4.1	-16.9	24.2	-23.1	18.8	-12.2
w	O-Si	354.8	Q5	-6.0	-5.6	-1.7	1.7	3.0	3.7	9.0	-5.2	-3.8	-13.8	27.5	-19.3	33.0	-9.1
x	C-S	395.8	Q3	1.4	6.8	5.0	3.3	3.3	2.8	2.8	6.8	5.9	16.3	-3.7	-16.1	1.4	12.8
y	S-S	390.3	Q4	-4.2	0.0	-1.7	-2.9	-2.9	-3.3	-3.3	-0.4	-1.0	-0.9	-9.7	-21.8	-4.2	-2.4
MUE				5.4	4.7	3.6	3.8	4.0	4.8	5.6	5.2	4.4	19.4	15.8	20.9	24.4	16.6

Table 2.6 Parameters of pseudopotentials for the tuned F link atoms

case	bond	CPS	bal. SEE	bal. RC	bal. RC2	Amber 2008-1	bal. RC3	Amber 2008-2	bal. RCD	bal. Shift
a	C–C	0.80	0.95	0.95	0.85	0.85	0.85	0.85	1.00	1.00
b	C–C	0.65	0.65	0.65	0.60	0.60	0.55	0.55	0.70	0.65
c	C–C	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.70	0.70
d	N–C	1.40	1.20	1.25	1.40	1.35	1.40	1.40	1.15	1.20
e	O–C	1.35	1.35	1.45	1.50	1.50	1.55	1.55	1.35	1.35
f	S–C	1.05	0.75	0.85	1.05	1.00	1.05	1.10	0.70	0.75
g	C–N	–0.20	–0.25	–0.15	0.05	0.00	0.10	0.15	–0.35	–0.35
h	O–N	1.00	0.35	0.60	0.85	0.85	0.95	0.95	0.30	0.45
I	C–O	–0.40	–0.30	–0.30	–0.25	–0.25	–0.25	–0.25	–0.30	–0.30
j	C–O	–0.25	–0.50	–0.45	–0.30	–0.30	–0.25	–0.25	–0.55	–0.45
k	C–O	–0.25	–0.50	–0.40	–0.25	–0.25	–0.20	–0.20	–0.55	–0.45
l	C–O	–0.45	–0.30	–0.30	–0.25	–0.25	–0.25	–0.25	–0.35	–0.30
m	Al–O	–0.15	–0.30	–0.20	–0.05	–0.05	–0.05	–0.05	–0.30	–0.25
n	Al–O	–0.10	–0.40	–0.20	–0.10	–0.10	–0.10	–0.10	–0.25	–0.15
o	Al–O	–0.15	–0.25	–0.20	–0.10	–0.10	–0.05	–0.05	–0.30	–0.25
p	Si–O	0.00	–0.25	–0.15	0.05	0.05	0.10	0.10	–0.30	–0.25
q	Si–O	–0.30	–0.45	–0.35	–0.25	–0.25	–0.20	–0.20	–0.45	–0.35
r	Si–O	0.00	–0.25	–0.15	0.05	0.05	0.05	0.10	–0.30	–0.15
s	C–Si	0.70	1.10	0.80	0.70	0.75	0.70	0.70	0.90	0.85
t	O–Si	0.60	0.80	0.60	0.60	0.60	0.60	0.60	0.70	0.60
u	O–Si	0.60	1.40	0.70	0.50	0.50	0.50	0.50	0.90	0.70
v	O–Si	0.65	1.40	0.80	0.65	0.65	0.60	0.55	1.00	0.80
w	O–Si	0.65	1.40	0.80	0.65	0.65	0.60	0.55	1.00	0.80
x	C–S	0.45	0.30	0.40	0.40	0.40	0.40	0.40	0.35	0.40
y	S–S	0.35	0.15	0.30	0.35	0.35	0.35	0.35	0.25	0.30

Table 2.7 Full-QM proton affinities (PA, in kcal/mol), QM/MM signed errors (in kcal/mol), and mean unsigned errors (MUE) (kcal/mol) of 25 cases by QM/MM calculations using tuned F link atoms

case	bond	PA	CPS	bal. SEE	bal. RC	bal. RC2	Amber-1	bal. RC3	Amber-2	bal. RCD	bal. Shift
a	C–C	362.3	2.6	–1.5	–0.8	0.0	–0.5	0.0	0.0	–2.0	–1.7
b	C–C	406.6	–6.2	–1.0	–0.2	0.6	–0.2	0.8	1.2	–1.1	–1.1
c	C–C	402.6	–2.0	1.6	2.0	3.1	2.8	3.5	3.5	1.5	1.6
d	N–C	376.7	0.1	1.7	1.0	–0.4	0.5	–1.3	–1.9	2.4	2.7
e	O–C	398.6	–0.2	0.1	–0.6	–1.4	–0.6	–1.7	–2.5	0.8	0.7
f	S–C	394.7	–0.8	1.1	0.3	–0.7	–0.2	–2.3	–2.5	1.5	1.2
g	C–N	355.2	2.8	4.0	2.5	–1.2	2.3	–5.6	–9.0	6.1	4.8
h	O–N	400.0	–3.4	4.6	2.8	–1.2	–1.3	–2.6	–2.6	6.3	5.3
I	C–O	394.9	0.6	1.0	1.1	0.4	0.4	–0.1	–0.5	2.0	1.8
j	C–O	401.0	–4.3	1.6	0.0	–2.5	–2.5	–3.6	–4.5	2.6	2.1
k	C–O	366.7	–1.6	2.0	0.2	–3.1	–3.1	–4.7	–5.7	3.4	2.5
l	C–O	398.3	–0.3	0.6	0.3	–0.1	–0.1	–0.4	–0.6	0.7	0.8
m	Al–O	340.7	1.7	1.8	0.3	–2.0	–2.0	–3.3	–3.3	2.8	1.8
n	Al–O	339.1	3.4	1.5	0.3	–2.3	–2.3	–3.6	–3.5	2.9	2.1
o	Al–O	348.0	–5.6	1.2	0.0	–1.9	–1.9	–2.8	–3.4	1.9	1.2
p	Si–O	349.0	–8.6	–1.2	–3.4	–6.5	–6.5	–7.9	–9.7	–0.2	–1.8
q	Si–O	353.2	–3.2	1.6	0.5	–1.1	–1.1	–1.9	–3.3	2.1	1.7
r	Si–O	348.1	–7.0	–0.3	–2.7	–6.1	–6.1	–7.9	–9.5	0.8	–0.6
s	C–Si	397.0	3.5	–5.0	–2.2	2.1	1.3	2.1	2.1	–6.5	–4.8
t	O–Si	342.7	–0.5	–2.2	–1.9	–0.5	–0.3	–0.5	–0.5	–2.7	–2.6
u	O–Si	348.0	–5.8	–2.7	–0.6	3.8	4.8	5.9	5.9	–5.1	–3.4
v	O–Si	353.2	–2.2	–0.7	0.4	3.1	4.2	4.7	6.9	–2.1	–1.4
w	O–Si	354.8	–3.8	–0.4	1.0	3.9	5.2	5.7	10.9	–1.8	–1.1
x	C–S	395.8	4.3	8.8	7.7	5.8	5.8	5.3	5.3	9.2	8.6
y	S–S	390.3	–1.9	1.0	0.4	–0.6	–0.6	–1.0	–1.0	1.3	1.0
	MUE		3.1	2.0	1.3	2.2	2.3	3.2	4.0	2.8	2.3

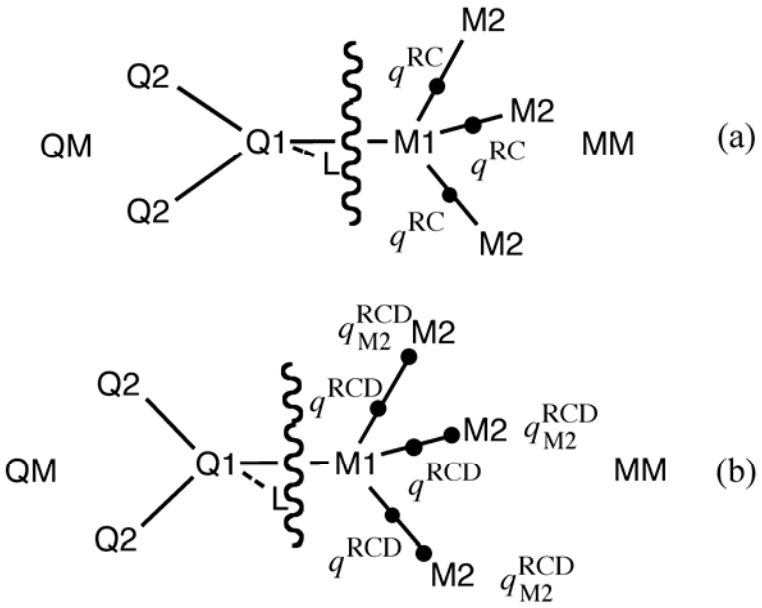


Figure 2.1 QM/MM boundary treatments in (a) the balanced RC scheme and (b) the balanced RCD scheme.

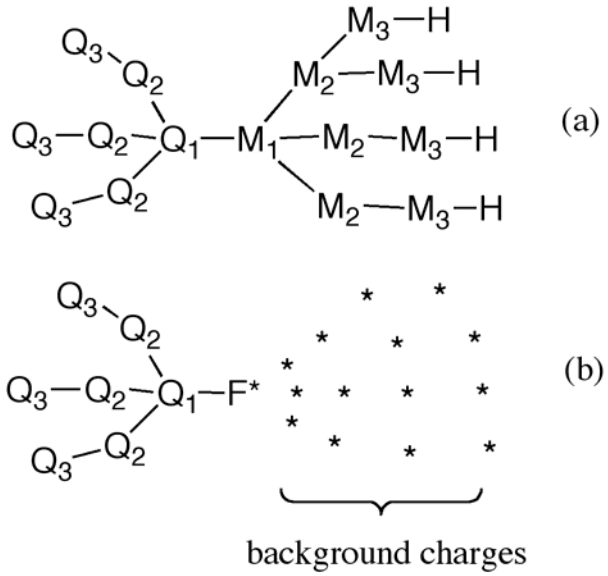


Figure 2.2 Determining the pseudopotential for the tuned F atom in the entire system model (ESM): (a) ESM and (b) CPS**, which is the capped QM subsystem with background charges to replace the rest of the ESM.

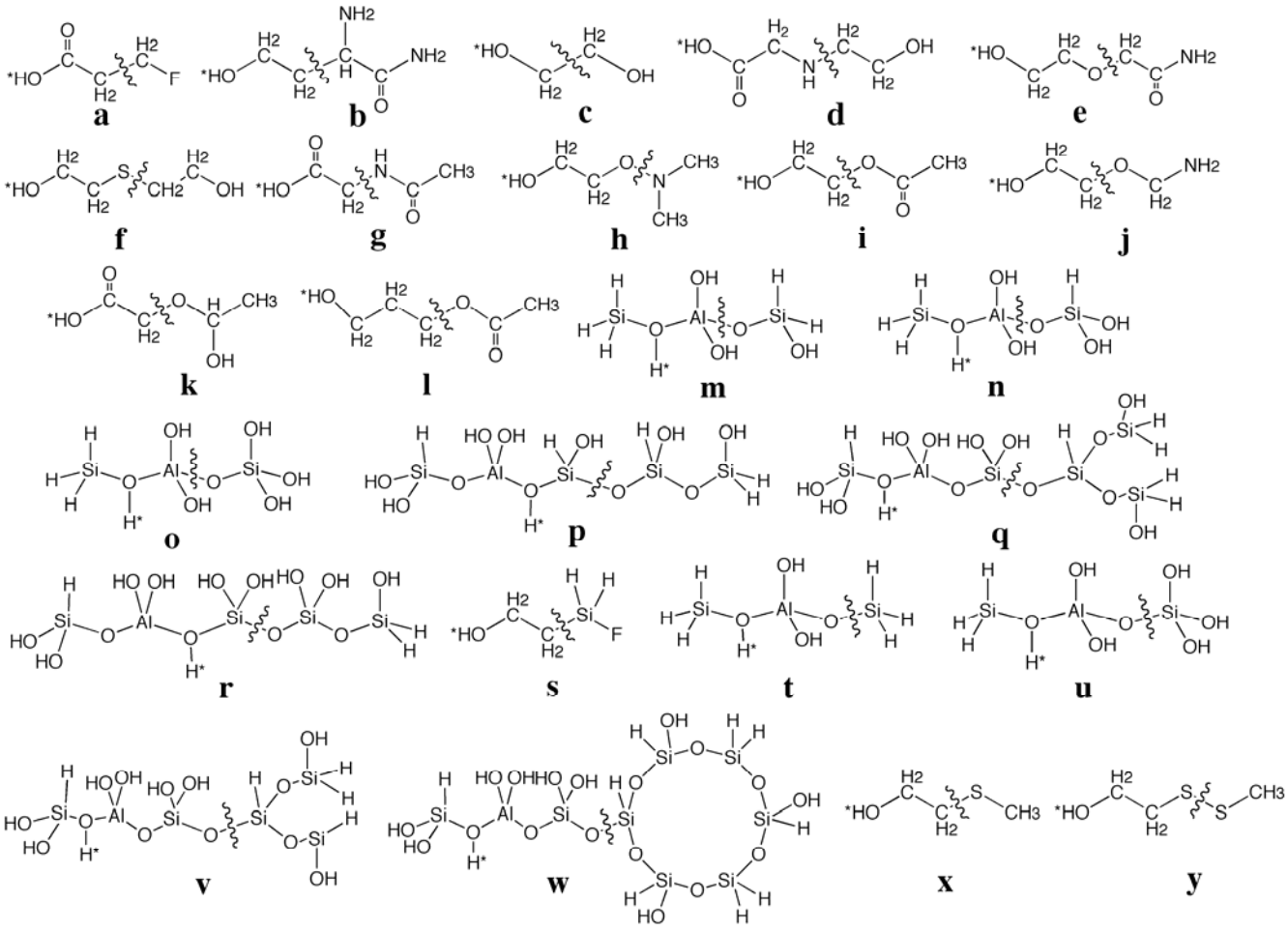


Figure 2.3 Twenty-five molecules used in the test suite. The QM subsystem is on the left, and the MM subsystem is on the right. * represents the proton involved in the protonation process.

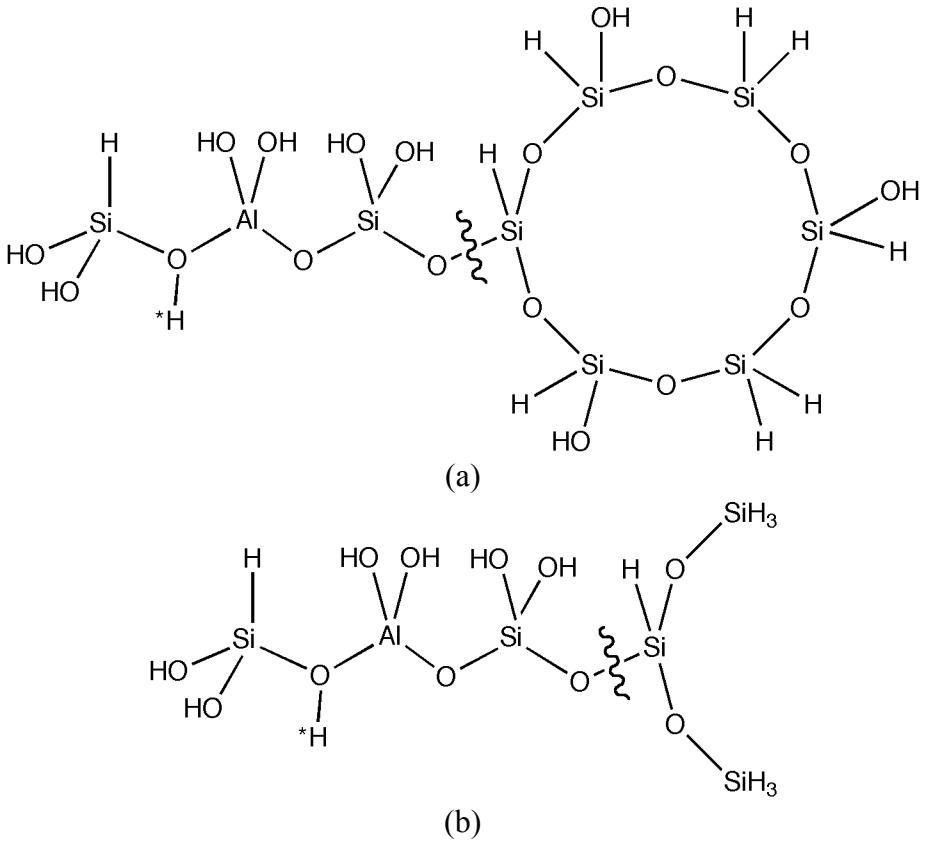


Figure 2.4 (a) The original entire system and (b) the entire system model (ESM) of the molecule **w**.

Chapter 3. Geometry Optimization Using Tuned and Balanced Redistributed Charge Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations[2]

3.1 Introduction

Multiscale modeling[1-6] is a method of choice for the study of chemical and physical processes of complex and large systems, such as practical catalysts and biomolecules. A key element is that a small-scale primary system is treated at a higher level than a large-scale secondary system, and there may even be a hierarchy of levels employed for more than two scales.

Combined quantum mechanical and molecular mechanical (QM/MM) methods are multiscale approaches that can be applied to study chemical reactions in large systems.[7-21] In QM/MM methods, a small region is treated by quantum mechanics, and the remaining part is treated by molecular mechanics. This method can be especially useful for simulations of condensed-phase systems, e.g., biomolecular processes and solid-state chemistry. For example, adsorption and chemical reactions in zeolites can be studied using QM/MM methods.[22]

One important issue in QM/MM methods is how to deal with the QM–MM boundary when it passes through a covalent bond. Link atoms,[7, 9, 10] generalized hybrid orbitals or other localized orbitals,[23-25] and pseudobonds or effective potentials[26-31] have been used to saturate the dangling valences at the edges of the QM region. To treat various kinds of

[2] This work was supported in part by the National Science Foundation under grant no. CHE09-56776.

covalent bonds, especially polar covalent bonds, being cut at the QM–MM boundary, we have developed tuned and balanced redistributed charge methods in a previous study; that article[32] will be called paper I. In these methods a pseudo atom tuned to reproduce the electronic structure of the QM region is used as the link atom, and the charge on the MM boundary atom is adjusted to conserve the total charge of the system[33] and is redistributed[34, 35] to other MM atoms. For example, in the tuned and balanced redistributed charge (TBRC) scheme, the adjusted (balanced) charge on the MM boundary atom is distributed to the midpoints of bonds between an MM boundary atom and its neighboring MM atoms. It is found that tuning the link atoms and adjusting (balancing) the MM charges can greatly improve the results for proton affinities in single-point calculations (i.e., calculations at fixed geometries), and the TBRC scheme gives the best results.[32]

In the present study, we formulate the QM/MM total energy expression as a function of geometry for systems with QM–MM boundaries that cut bonds, and we carry out QM/MM geometry optimizations. We also add a new option to the tuned and balanced redistributed charge methods, namely charge smearing. When this option is employed, the redistributed charges are not represented by point charges, but by smeared charges. Das et al.[36] and Amara and Field[37] have represented the charge distributions of MM atoms close to a QM region by Gaussian functions rather than point charges. For boundaries that cut a C–C bond, they showed that when the nearby MM charges are properly delocalized, the description of electrostatics nears the QM–MM boundary is improved. We have developed[38] a particularly convenient way to delocalize the outermost portion of an atom's charge in a Slater-type orbital[39] (STO). In the present

work, we use a similar scheme to delocalize the redistributed charges near a cut covalent bond.

A key element to be examined is whether the tuned and balanced redistributed charge schemes improve the accuracy enough to optimize geometries realistically and calculate reaction energies accurately with optimized geometries. This is important not only for geometry optimization *per se* but for molecular dynamics; a method that yields incorrect geometries when both lengths and bond angles are unconstrained is not suitable for molecular dynamics. Here we test the tuned and balanced redistributed charge schemes for this capability using both the point charges and smeared charges for the balanced redistributed charges. In section 3.2, we will present all the ingredients of the QM/MM methods used in the present study. In section 3.3, we will present the test suite and implementation details. Section 3.4 gives the analysis of the calculations. Section 3.5 gives an overall comparison of the performance of all boundary charge schemes. Section 3.6 summarizes the main conclusions.

3.2 Methods

In this section, we will first review the tuned and balanced redistributed charge schemes proposed in paper I.[32] Then we will present the QM/MM energy expression, which is based on an earlier[35] formulation. This is followed by a description of the placement of the link atom and the smearing of the redistributed charges.

In order to describe the schemes, we label the atoms according to "tiers".[12, 32, 35] In particular, the MM boundary atoms (i.e., MM atoms covalently bonded to QM atoms) are denoted as M1 atoms; and MM atoms directly bonded to M1 atoms are denoted as M2 atoms. M3 atoms are the third-tier MM atoms, i.e., those bonded to M2 atoms. The QM

boundary atoms (i.e., QM atoms covalently bonded to MM atoms) are denoted as Q1 atoms; and the QM atoms directly bonded to Q1 atoms are labeled Q2 atoms.

3.2.1 Tuned and Balanced Redistributed Charge Schemes

In paper I,[32] we introduced tuned and balanced redistributed charge schemes to treat the QM–MM boundary. The balancing consists in adjusting the MM point charge on the M1 atom to conserve the total charge of the entire QM/MM system. In the BRC scheme, this adjusted charge is evenly redistributed to the midpoints of the M1–M2 bonds. The adjusted M1 charge can also be placed in other positions. In particular, when the adjusted M1 charge is evenly redistributed to all M2 atoms, we call the method BRC2, and when the adjusted M1 charge is evenly redistributed to all M2 and M3 atoms, we call the method BRC3. In the balanced redistributed charge and dipole (BRCD) scheme, we double the redistributed charges that are placed at the midpoints of the M1–M2 bonds, and we adjust the charges on M2 atoms to conserve the total charge of the entire QM/MM system.[32] To test the electrostatic effects of the MM charges on the QM region, we also test a Z∞ scheme, in which the electrostatic interactions between the QM and MM regions are completely neglected.

In tuned methods, the link atom is a tuned F atom, which is an atom that has an adjustable pseudopotential centered at its nucleus. The pseudopotential is given by

$$U(r) = C\exp[-(r/r_{\mathrm{T}})^2] \tag{3.1}$$

where C is the tuning parameter, and $r_{\mathrm{T}} = 1\,a_0$ (where a_0 is Bohr radius. The pseudopotential is tuned to make the sum of the partial charges of the uncapped portion of the QM subsystem equal a target value. The partial charges are computed by Mulliken

analysis with a 6-31G* basis set when the M1 atom is from the second period (Li through F) and with an STO-3G basis set otherwise. The tuning process has been used successfully[32] to treat polar bonds between the QM and MM subsystems. A detailed description of the tuning process can be found in paper I. We combine the tuned F link scheme with the BRC, BRC2, BRC3, and BRCD schemes, yielding the TBRC, TBRC2, TBRC3, and TBRCD schemes, respectively.

In paper I we tuned the fluorine link atoms on an entire system model (ESM) that includes three tiers of MM atoms, with the third tier capped. Background charges in the ESM (i.e., MM charges and redistributed charges) are present in the tuning process. In the present study, we found that tuning the F link atoms with and without background charges gives similar tuning parameters. Tuning without background charges is more straightforward, so in the present work, we tuned the F link atom without any background charges present, which, in the language of paper I, corresponds to tuning the capped primary system (CPS) without any MM charges. A consequence of this simplification that is important for the present work is that the tuning parameters are independent of the charge models and smearing widths that are used below to treat MM charges and boundary charges.

3.2.2 Total energy expression

The QM region is also called the primary subsystem (PS), the MM region is called the secondary subsystem (SS), and the whole system is called the entire system (ES). The CPS is the primary system (PS) capped by the link atom. CPS** denotes that the CPS is embedded in the adjusted electrostatic field of the secondary subsystem (or the unadjusted one in the straight electrostatic embedding (SEE) method), and SS* denotes

the secondary subsystem with the adjusted M1 charge in the balanced charge methods or with the original M1 charge in the unbalanced charge methods. No charge redistribution is made for SS*. The QM/MM energy is[35]

$$E = E(\mathrm{QM,CPS**}) + [E(\mathrm{val;ES}) - E(\mathrm{val;CPS})] + E(\mathrm{Coul;SS*}) + [E(\mathrm{vdW;ES}) - E(\mathrm{vdW;CPS})] \quad (3.2)$$

where $E(\mathrm{QM,CPS**})$ is the quantum mechanical energy of the QM system in the presence of (original or adjusted) electrostatic field of the secondary subsystem, the two differences, $E(\mathrm{val;ES}) - E(\mathrm{val;CPS})$ and $E(\mathrm{vdW;ES}) - E(\mathrm{vdW;CPS})$, are the MM energy differences between the entire system and the capped primary system for the valence interactions and van der Waals interactions, and $E(\mathrm{Coul;SS*})$ is the Coulomb energy of the secondary subsystem with the original or adjusted M1 charge. Therefore, we extend the energy expression formulated in the RC and RCD article[35] to the new balanced charge schemes, such as BRC and BRC2. The only difference between the BRC scheme and the original RC scheme[35] is that the adjusted (balanced) M1 charge, rather than the original M1 charge, is used for the calculations of $E(\mathrm{QM,CPS**})$ and $E(\mathrm{Coul;SS*})$. In the special case where the total charge of the MM region is neutral, the adjusted M1 charge equals the original M1 charge, and these two formulations are same.

3.2.3 Placement of the link atom

The tuned F link atom is not at its equilibrium position in the QM/MM calculation unless geometry optimization is performed. In the current study, we used the same approach as that used in *Amber 10*[33] to place the link atom (which is denoted as L); in this method the Q1–L bond length is fixed during the QM/MM optimization. The link atom is

placed along the bond vector joining Q1 and M1, and the position of the link atom is defined as:

$$\mathbf{r}_{\mathrm{L}} = \mathbf{r}_{\mathrm{Q1}} + d_{\mathrm{Q1\text{-}L}} \frac{\mathbf{r}_{\mathrm{M1}} - \mathbf{r}_{\mathrm{Q1}}}{\left|\mathbf{r}_{\mathrm{M1}} - \mathbf{r}_{\mathrm{Q1}}\right|} \tag{3.3}$$

where $\mathbf{r}_{\mathrm{L}}$, $\mathbf{r}_{\mathrm{Q1}}$, and $\mathbf{r}_{\mathrm{M1}}$ are the positions of the link atom, the Q1 atom, and the M1 atom, respectively; and $d_{\mathrm{Q1?L}}$ is the fixed bond length of Q1–L, which is assigned as the standard Q1–L bond length in whatever force field is used for the MM calculations.

3.2.4 Smearing the redistributed charges

For the option of charge smearing, we placed the redistributed charge q_{MM} in a normalized Slater type orbital (STO)

$$\varphi = \exp(-r / r_0) / (\pi r_0^{\,3})^{1/2} \tag{3.4}$$

where r is the distance of the charge density from its center, and r_0 is the smearing width. Then the charge density of MM charge q_{MM} is expressed as

$$\rho_{\mathrm{MM}}(r) = q_{\mathrm{MM}} \exp(-2r / r_0) / (\pi r_0^{\,3}) \tag{3.5}$$

We calculated the electrostatic potential generated by the smeared charge and derived the effective charge as

$$q^{*}_{\mathrm{MM}} = q_{\mathrm{MM}} - q_{\mathrm{MM}}(1 + \frac{r}{r_0}) \exp(-2r / r_0) \tag{3.6}$$

An explanation of the effective charge concept and a detailed derivation can be found in our previous paper about charge penetration.[38] The only difference here is that we delocalized the outermost electron density in the study of charge penetration,[38] while we delocalize the total redistributed charges in the present study. Because in most programs,

the pseudopotentials are expressed as Gaussian types of functions, we used 6 Gaussian functions to fit the Slater function, as is shown in eq 3.7, and the contraction coefficients C_i and exponents α_i are listed in Table 3.1.[40]

$$\exp(-\lambda r) = \sum_{i=1}^{6} C_i \exp(-\alpha_i \lambda^2 r^2) \tag{3.7}$$

The reason we use this particular way to smear the MM charges is that these effective charges can be easily implemented as pseudopotentials in standard QM programs.

In the present study, we only smear the redistributed charges, while point charges are still used for all other MM charges. Combining the smeared redistributed charge (SRC) scheme with the TBRC, TBRC2, TBRC3, and TBRCD schemes yields the TBSRC, TBSRC2, TBSRC3, and TBSRCD schemes.

3.3 Computational details

All computations were carried out by using the M06-2X density functional method[41, 42] as the QM method. The MMFF94 force field[43] was used for the *E*(val) and *E*(vdW) terms of eq 3.2. For the MM charges, M06-2X/6-31G(d)/CM4M[44] charges are used because they seem to be reasonable choices to reproduce the electrostatic potentials generated by MM atoms, and they are more accurate for the buried atoms in the systems studied here than are the CHELPG charges. The CM4M charges are derived from the protonated molecules, and–as in the usual procedure in universal MM force fields–are assumed to be the same in the unprotonated molecules as in the protonated ones. The MM parameters for the tuned F atom are taken to be same as that for the ordinary F atom.

All QM/MM calculations are carried out using our own QMMM program,[45] which is based on a locally modified module[46] of *Gaussian03*[47] and a modified TINKER[48] program.

M06-2X/6-31G(d)/CM4M[44] charges are derived from a locally modified module[49] of *Gaussian03*.[47] In the current study, three basis sets are tested, in particular 6-31G(d),[50-52] def2-TZVP,[53] and MG3S.[54] The 6-31G(d) basis set is used for all calculations in sections 3.4.1–3.4.3, and the other basis sets are considered in section 3.4.4.

Both the protonated and unprotonated molecules are optimized using both full QM and combined QM/MM methods. We compared the QM/MM deprotonation energies and geometries to the QM results for a test suite that contains 15 molecules with 10 kinds of single bonds being cut, in particular C–C, C–N, C–O, O–C, N–C, C–S, S–S, S–C, C–Si, and O–N bonds. The protonated forms of molecules in the test suite are shown in Figure 3.1. Because MMFF94 does not include all MM parameters that are required to treat aluminosilicate clusters and we need to examine the validity of available aluminosilicate force fields in a QM/MM context before using them to test the new methods discussed here, we exclude the aluminosilicate clusters that were tested in paper I.

For the geometries, we considered the Q1–M1, M1–M2, and Q1–Q2 bond distances for the QM and QM/MM optimized structures. If there is more than one M1–M2 or Q1–Q2 bond, only the bond not involving a hydrogen atom is counted. For M1–M2 bonds in molecule **ON_1**, there are two N–C bonds and only the longer one is included in the error analysis.

3.4 Results and discussion

In this section, we first present the tuning parameters used for the tuned F link atom in the present study. This is followed by the QM/MM results using H link atoms and tuned F link atoms with the redistributed charges described by point charges. Then we show the QM/MM results with the redistributed charges being smeared. The smearing widths of

the redistributed charges have been optimized for several tuned and balanced redistributed charge schemes. Finally we tested the transferability of the tuning parameters and smearing widths in the TBRC2 and TBSRC schemes to other basis sets.

3.4.1 Tuning parameters for the tuned F link atom

We performed the tuning process for each protonated molecule in the test suite. Table 3.2 shows the parameters *C* of the tuned F link atoms in eq 3.1 for all molecules. Because the tuned F link atoms are tuned without the background charges (as explained in Sect. 3.2.1), the same parameters are used for all charge models and all smearing widths.

Note that the parameters we used here are different from the ones we used in paper I, in which the parameters are tuned in the presence of three tiers of MM charges.[32] Also, the Q1–L bond length is taken in the present study as the standard bond length of Q1–F in the MMFF94 force field, which also differs from paper I. Different placements of the link atom cause some differences in the values of the tuning parameters, but the differences are relatively small. Nevertheless, all results in the present article are recalculated in the new way explained above.

3.4.2 Redistributed charges as point charges

An H atom has been used as the link atom in most previous QM/MM methods, so we used both H atoms and tuned F atoms as link atoms for the QM/MM calculations. We carried out the QM/MM optimization using the BRC, BRC2, BRC3, BRCD, TBRC, TBRC2, TBRC3, and TBRCD schemes. To make a comparison, we also show the results using the Z∞ scheme. In the Z∞ scheme, all MM charges are zeroed for the calculations

of E(QM,CPS**) in eq 3.2, and other terms in eq 3.2 are evaluated in the same way as in the BRC scheme.

Table 3.3 shows the results for various charge schemes using H link atoms and tuned F link atoms. When H atoms are used as link atoms, all schemes, including the BRC, BRC2, BRC3, and BRCD schemes, give quite accurate geometries. This indicates that in these redistributed charge schemes with H link atoms, the redistributed charges are already moved far enough from the QM–MM boundary to avoid overpolarization. However, the mean unsigned error (MUE) of deprotonation energies is 6.6–9.2 kcal/mol for various schemes (Z∞, BRC, BRC2, BRC3, and BRCD) employing H link atoms.

When tuned F atoms are used as link atoms, the MUE of deprotonation energies drop to 1.6–3.6 kcal/mol for the Z∞, TBRC2, TBRC3 schemes, with the best performance from the TBRC2 scheme. At the same time, the geometries are also well reproduced in QM/MM optimizations. However, for the TBRC and TBRCD schemes, we found that nearly half of the molecules are severely distorted and are not converged in QM/MM optimizations, and we do not list their errors. One possible reason for the unphysical behavior is that in these two schemes, the redistributed charges located at the midpoints of the M1–M2 bonds are close to the tuned F link atom and may overpolarize the QM region. Hence the TBRC method as originally formulated overestimates induction energies; nevertheless, the errors are largely cancelled in computing relative energies by single-point calculations, and so we obtained good relative energies in the previous study. However, the strong interactions between the redistributed charges and the link atom distort the structures in QM/MM geometry optimizations.

3.4.3 Redistributed charges as smeared charges

To overcome the problem due to the strong electrostatic interactions between the QM region and the redistributed charges in the TBRC and TBRCD schemes, we used smeared charges to represent the redistributed charges, leading to the TBSRC and TBSRCD schemes. It has been found that smearing the MM charges close to the QM region reduces the large electrostatic interactions near the boundary.[36] We also tested the TBSRC and TBSRC3 schemes, in which the smeared redistributed charges (SRC) are combined, to study the effects of charge smearing, though the TBRC2 and TBRC3 schemes give reasonable optimized geometries.

We first tested the TBSRC and TBSRC2 schemes plus two redistributed charge schemes in which the positions of the redistributed charges are varied. The parameter F is defined as the ratio

$$F = \frac{R(\mathrm{M1-RC})}{R(\mathrm{M1-M2})} \tag{3.8}$$

where $R(\mathrm{M1-RC})$ is the distance from M1 atom to the redistributed charge, and $R(\mathrm{M1-M2})$ is the distance from M1 atom to M2 atom. The F value is 0.5 for TBSRC, and it is 1.0 for TBSRC2. The other two schemes have F values of 0.3 and 0.7.

In Tables 3.4 and 3.5, we show the MSE and MUE of the bond lengths and deprotonation energies using tuned F link atoms with various smearing widths and with four different positions for the redistributed charges. We found that the QM/MM optimization is successful in nearly all cases except one. Table 3.4 shows that when the smearing width increases, the MUE of the bond lengths decreases. This confirms that smearing the redistributed charges can improve the electrostatics nears the QM–MM boundary and give better optimized geometries. The smearing width needs to be large

enough to avoid the distortion of QM/MM structures. If *F* equals 0.3 or 0.5, the smearing width needs to be greater or equal 1.0 Å. If *F* equals 0.7, the smearing width needs to be greater or equal to 0.5. If *F* equals 1.0, there is no need to smear the charges. As the redistributed charges are moved farther from the QM/MM boundary, the overpolarization problem is less severe, and we can use smaller smearing width to get correct geometries. At the same time, Table 3.5 shows that when the smearing width becomes larger, the MUE of the deprotonation energy increases in most cases, especially for TBSRC2. In the special case in which the smearing width goes to infinity, QM atoms will not feel the redistributed charges and the total charge of the QM/MM entire system is not conserved, which will cause large errors.[32] The TBSRC scheme (*F* equals 0.5) with the smearing width of 1.0 Å gives the smallest MUE, 1.62 kcal/mol, for the deprotonation energy.

The QM/MM results using the TBSRC3 and TBSRCD schemes are shown in Table 3.6. For the TBSRC3 scheme, the MUE of the deprotonation energies increases when we increase the smearing width of the redistributed charges. For the TBRCD scheme, the optimum value of the smearing width is 2.0 Å, with the MUE of deprotonation energies of 1.76 kcal/mol.

From the above results, we conclude that for the TBRC and TBRCD schemes, it is necessary to include the smeared redistributed charge (SRC) scheme to reproduce both the geometries and deprotonation energies. For the TBRC2 and TBRC3 schemes, in which the redistributed charges are moved farther from the QM–MM boundary, charge smearing is not necessary and the schemes without charge smearing give the best results.

We also tested whether smearing the redistributed charges with H link atoms can give good results for this test suite. Table 3.7 shows the results using H link atoms with

various smearing widths. Varying the smearing widths from 0.0 to 5.0 Å can change the MSE of the deprotonation energies by up to 4 kcal/mol, and a smearing width of 3.0 Å gives the smallest MUE of the deprotonation energy of 6.7 kcal/mol. In previous studies,[36, 37] it was shown that properly adjusting the smearing width of smeared MM charges can greatly reduce the MUE of protonation and deprotonation energies when C–C bonds being cut in the boundary. However, when polar bonds are cut in the QM–MM boundary, as in the present work and paper I, H link atoms may change the electronic structure of the boundary in the QM region, and smearing the redistributed charges does not correct the error.

3.4.4 Tests with other basis sets

To test the transferability of the scheme among various QM basis sets, we carried out calculations with two other basis sets, in particular def2-TZVP and MG3S. The M06-2X/6-31G(d)/CM4M charges are again used as MM charges. Two methods selected from Sect. 3.4.2 and 4.3 have been tested: one is the TBRC2 scheme, and the other one is the TBSRC scheme with a smearing width of 1.0 Å for redistributed charges. The results are shown in Table 3.8. Although for both the TBRC2 and TBSRC schemes, the MUEs of the deprotonation energies, which are around 2.3 kcal/mol, are slightly larger than were found using the 6-31G(d) basis set, they are still much smaller than their counterparts using H link atoms (H link atoms give MUEs of 6–7 kcal/mol). Therefore we conclude that our scheme is transferable, to some extent, among different basis sets. Moreover, we found that def2-TZVP and MG3S, which are more complete basis sets than 6-31G(d), have significant mean signed errors (MSE) with the deprotonation energies in both cases.

The systematic character of the error indicates that further developments may be able to reduce the error for large QM basis sets.

3.5 Comparison of methods

Table 3.9 provides a consistent overall comparison of the performance of all considered boundary charge schemes for calculating proton affinities of the 15-molecule test suite considered here. The table has eight numerical columns; the first four refer to the entire-system geometries optimized in paper I.[32] The next four refer to geometries optimized by QM/MM calculations in the present work. All QM calculations, including the QM parts of the QM/MM calculations are based on M06-2X/6-31G(d). The MM force field is MMFF94 except that we use CM4M partial atomic charges. The first three columns of the table are computed for the present 15-molecule test set from calculations originally carried out for paper I. The next five columns are based on new calculations carried out for the present study.

In order to remind the reader of the differences among the redistributed charge schemes, we added an extra column for those methods to show the sites that are affected by charge redistribution. M1, M2 and M3 sites have already been defined; M1.5 denotes a site halfway between M1 and M2; M2D denotes added dipole sites near M2. The shift method is due to Sherwood et al;[34] it is similar to RCD, but we have preferred RCD because RCD is less complicated. The Amber-1 method[33] is the *adjust_q=1* method in the *AMBER 10* program,[55] and Amber-2 is the default method in that program. Amber-1 is similar to RC2 but has the disadvantage that the redistributed charge can hop discontinuously among M2 atoms as the M1–M2 bonds vibrate. Amber-2 is similar to

RC3, but we consider Amber-2 to be unphysical because it redistributes charges to arbitrarily distant locations. Nevertheless, we show these results for comparison.

The first row of Table 3.9 shows results with the MM subsystem neglected. This provides a baseline that methods including the MM subsystem should surpass. The effect tuning at this level is remarkable. The MUE is only 2.3 kcal/mol for single-point calculations and 3.5 kcal/mol for optimized geometries; these results are better than any results (5–19 kcal/mol) obtainable with hydrogen link atoms or untuned fluorine link atoms. This shows that attempts to improve untuned link atom methods by improving the boundary charges are missing the point; the dominant error in untuned link atom methods is the incorrect charge distribution in the QM subsystem itself, not in its interaction with the MM subsystem.

Turning our attention to methods including the MM subsystem, we first consider the unbalanced methods. SEE denotes straight electrostatic embedding (the partial charge on M1 is not redistributed), Z2 denotes the default of the ONIOM method as implemented in the *Gaussian 03*[47] and *Gaussian 09*[56] packages in which the charges on the first two tiers of MM atoms are just set equal to zero. We see error for these methods in the range 17–19 kcal/mol. The RC scheme lowers these errors only to 16 kcal/mol. To do better we must use balanced schemes and tuning schemes.

The third and fourth numerical columns of Table 3.9 compare the two ways to conduct the tuning. The original method (see paper I) for tuning involved tuning in the presence of MM charges in a three-tier entire system model (ESM). The simpler method, introduced in the present paper, is to tune in the absence of MM charges. On average, tuning in the presence of MM charges in the ESM lowers the MUE by 2.9 kcal/mol,

whereas tuning in the CPS lowers the error by 2.5 kcal/mol. These lowerings are close enough to one another that we selected to use the much more straightforward CPS tuning in the rest of the work, which is shown in the last two columns of Table 3.9.

Finally, we consider methods that employ both balanced charges and tuning. First consider balanced straight electrostatic embedding (BSEE). Here the balanced charges on M1 stay on M1. Table 3.9 shows that geometry optimization with this method is only successful if we employ smearing with H link atoms (the result shown is for a smearing width of 1.0 Å), and even when the geometry optimization is converged, the error is large. We conclude that some redistribution is required.

Second, the results show that charge smearing is necessary if the redistributed charge is close to the QM region (as in the TBSRC and TBSRCD), but not if it is farther from the QM region (as in the TBRC2 and TBRC3 methods).

Considering the last two columns of the last four rows of Table 3.9, we see that methods employing the M1.5 site are suitable for geometry optimizations only if charge smearing is employed, but when this is done, the MUE drops to 1.6 kcal/mol. The overall best method though seems to be TBRC2. It is simpler than TBRC or TBRCD in that it does not involve the M1.5 site, and it does not require charge smearing; and it is one of the most accurate methods in Table 3.9 when tuning is based on the CPS, except perhaps for Amber-1, which we did not pursue because of the discontinuous charge redistribution problem.

3.6 Conclusions

QM/MM optimizations have been performed using the tuned and balanced redistributed charge schemes. A charge-smearing scheme for the redistributed charges

has been introduced in order to make the electrostatic interactions near the QM–MM boundary more realistic. It is found that both QM/MM optimized geometries and QM/MM deprotonation energies calculated with optimized geometries can accurately reproduce full QM results even for boundaries through polar bonds, and there are also significant improvements for boundaries through C–C bonds. Both the TBRC2 scheme and the TBSRC scheme with a smearing width of 1.0 Å give a mean unsigned error of 1.6 kcal/mol for the deprotonation energies, and the QM/MM optimized geometries also agree well with the QM geometries for these two choices. Moreover, comparing the results using H link atoms to those tuned F link atoms, we conclude that it is necessary to tune the link atoms when treating diverse kinds of bonds at the QM–MM boundary. In fact tuning and balancing are found to be more important than the choice of charge redistribution scheme, although the literature devoted to tuning and balancing is small, and that devoted to charge redistribution is large.

3.7 References for Chapter 3

1 G. S. Ayton, W. G. Noid and G. A. Voth, *Curr. Opin. Struct. Biol.*, 2007, **17**, 192.
2 A. Heyden, H. Lin and D. G. Truhlar, *J. Phys. Chem. B*, 2007, **111**, 2231.
3 M. Praprotnik, L. D. Site and K. Kremer, *Annu. Rev. Phys. Chem.*, 2008, **59**, 545.
4 P. Sherwood, B. R. Brooks and M. S. P. Sansom, *Curr. Opin. Struct. Biol.*, 2008, **18**, 630.
5 T. Murtola, A. Bunker, I. Vattulainen, M. Deserno and M. Karttunen, *Phys. Chem. Chem. Phys*, 2009, **11**, 1869.
6 M. Tafipolsky, S. Amirjalayer and R. Schmid, *Micropor. Mesopor. Mater.*, 2010, **129**, 304.
7 U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, 1986, **7**, 718.
8 J. Gao, *Rev. Comp. Chem*, 1996, **7**, 119.
9 D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580.
10 I. Antes and W. Thiel, *J. Phys. Chem. A*, 1999, **103**, 9290.
11 N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720.
12 P. Sherwood, *Modern Methods and Algorithms of Quantum Chemistry*, John von Neumann Institute for Computing: Jülich, 2000, pp 285.

13 J. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467.
14 A. Shurki and A. Warshel, in *Protein Simulations*, Academic Press Inc, San Diego: USA, 2003, vol. 66, pp. 249.
15 R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389.
16 M.-E. Moret, E. Tapavicza, L. Guidoni, U. Röhrig, M. Sulpizi, I. Tavernelli and U. Rothlisberger, *Chimia*, 2005, **59**, 493.
17 T. Vreven and K. Morokuma, *Annu. Reports Comp. Chem.*, 2006, **2**, 35.
18 H. M. Senn and W. Thiel, *Curr. Opin. Chem. Biol.*, 2007, **11**, 182.
19 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185.
20 H. Hu and W. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573.
21 H. M. Senn and W. Thiel, *Angew. Chemie Int. Ed.*, 2009, **48**, 1198.
22 B. Boekfa, S. Choomwattana, P. Khongpracha and J. Limtrakul, *Langmuir*, 2009, **25**, 12990.
23 V. Théry, D. Rinaldi, J. L. Rivail, B. Maigret and G. G. Ferenczy, *J. Comp. Chem.*, 1994, **15**, 269.
24 J. Gao, P. Amara, C. Alhambra and M. J. Field, *J. Phys. Chem. A*, 1998, **102**, 4714.
25 J. Pu, J. Gao and D. G. Truhlar, *J. Phys. Chem. A*, 2004, **108**, 632.
26 N. Koga and K. Morokuma, *Chem. Phys. Lett.*, 1990, **172**, 243.
27 Y. Zhang, T.-S. Lee and W. Yang, *J. Chem. Phys.*, 1999, **110**, 46.
28 F. Alary, R. Poteau, J. L. Heully, J. C. Barthelat and J. P. Daudey, *Theor. Chem. Acc.*, 2000, **104**, 174.
29 G. A. DiLabio, M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.*, 2002, **116**, 9578.
30 O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *J. Chem. Phys.*, 2005, **122**, 014113.
31 Y. Y. Ohnishi, Y. Nakao, H. Sato and S. Sakaki, *J. Phys. Chem. A*, 2008, **112**, 1946.
32 B. Wang and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 359.
33 R. C. Walker, M. F. Crowley and D. A. Case, *J. Comp. Chem.*, 2008, **29**, 1019.
34 P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent and I. H. Hillier, *Faraday Discuss.*, 1997, **106**, 79.
35 H. Lin and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 3991.
36 D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček and B. R. Brooks, *J. Chem. Phys.*, 2002, **117**, 10534.
37 P. Amara and M. J. Field, *Theor. Chem. Acc.*, 2003, **109**, 43.
38 B. Wang and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 3330.
39 J. C. Slater, *Phys. Rev.*, 1930, **36**, 57.
40 W. J. Hehre, R. F. Stewart and J. A. Pople, *J. Chem. Phys.*, 1969, **51**, 2657.
41 Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157.
42 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.
43 T. A. Halgren, *J. Comp. Chem.*, 1996, **17**, 490.
44 R. M. Olson, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 2046.
45 H. Lin, Y. Zhang and D. G. Truhlar, QMMM, version 1.3.6, University of Minnesota: Minneapolis, 2009.
46 Y. Zhao and D. G. Truhlar, MN-GFM: Minnesota Gaussian Functional Module, version 3.0, University of Minnesota: Minneapolis, 2006.

47 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03*, version D. 01, Gaussian Inc.: Wallingford, CT, 2004.
48 J. W. Ponder, TINKER, version 5.0, Washington University: St. Louis, 2009.
49 A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, Minnesota Solvation Database, version 2008, University of Minnesota: Minneapolis, 2008.
50 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257.
51 J. D. Dill and J. A. Pople, *J. Chem. Phys.*, 1975, **62**, 2921.
52 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654.
53 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys*, 2005, **7**, 3297.
54 B. J. Lynch, Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 1384.
55 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 10*, University of California: San Francisco, 2008.
56 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 09*, version A. 02, Gaussian Inc.: Wallingford CT, 2009.

Table 3.1 The contraction coefficients and exponents[a]

	C_i	α_i
1	0.021221	0.065110
2	0.131906	0.158088
3	0.238573	0.407099
4	0.241818	1.185060
5	0.184113	4.235920
6	0.121984	23.10300

[a] From ref. 40

Table 3.2 Parameters of pseudopotentials for the tuned F link atoms

Molecule	**CO_1**	**CO_2**	**CO_3**	**CO_4**	**CN_1**	**CC_1**	**CC_2**	**CC_3**
Parameter	–0.40	–0.15	–0.10	–0.40	–0.05	0.90	0.70	0.75
Molecule	**NC_1**	**OC_1**	**CS_1**	**SS_1**	**SC_1**	**CSi_1**	**ON_1**	
Parameter	1.40	1.35	0.45	0.40	1.00	0.70	1.05	

Table 3.3 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM bond lengths (Å) and deprotonation energies (kcal/mol) using H link atoms and tuned F link atoms with point charge representation of the redistributed charge.

Link atom	Charge scheme	bond length[a] (Å)		deprotonation energy (kcal/mol)	
		MSE	MUE	MSE	MUE
H atom	Z∞	0.003	0.014	8.48	8.58
	BRC	–0.002	0.016	8.30	8.30
	BRC2	0.004	0.014	7.27	7.41
	BRC3	0.004	0.014	6.22	6.55
	BRCD	–0.010	0.021	9.20	9.20
Tuned F atom	Z∞	–0.002	0.016	1.24	3.59
	TBRC2	–0.002	0.015	0.48	1.65
	TBRC3	–0.002	0.015	–0.55	2.44

[a] Each mean error in bond length is an average over 90 values: 15 molecules, each protonated and unprotonated, and three bond distances for each protonated or unprotonated molecule.

Table 3.4 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM bond lengths (Å) using tuned F link atoms with various positions of the smeared redistributed charge and with various smearing widths r_0 (Å).[a]

F	0.3		0.5		0.7		1.0	
r_0(Å)	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
0.5	–0.022[b]	0.070[b]	–0.025	0.048	–0.018	0.026	–0.007	0.014
1.0	–0.008	0.028	–0.010	0.021	–0.010	0.017	–0.007	0.014
2.0	–0.008	0.016	–0.005	0.014	–0.005	0.014	–0.005	0.014
3.0	–0.002	0.016	–0.005	0.014	–0.003	0.014	–0.004	0.014

[a] TBSRC for $F = 0.5$ and TBSRC2 for $F = 1.0$. The other columns do not have a name but may be considered to be nonstandard variants of TBSRC.

[b] When $r_0 = 0.5$ Å and $F = 0.3$, optimization of **CO_3** (unprotonated form) is not converged, so it has been excluded from calculations in that case.

Table 3.5 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM deprotonation energies (kcal/mol) using tuned F link atoms with various positions of the redistributed charge and with various smearing widths r_0 (Å).[a]

F	0.3		0.5		0.7		1.0	
r_0(Å)	MSE	MUE	MSE	MSE	MSE	MUE	MSE	MUE
0.5	1.50[b]	2.30[b]	1.22	1.93	0.87	1.63	0.35	1.80
1.0	1.16	1.73	0.82	1.62	0.52	1.66	0.09	2.10
2.0	0.10	2.18	−0.05	2.38	−0.26	2.64	−0.55	3.01
3.0	−0.71	3.42	−0.91	3.65	−1.01	3.81	−1.21	4.09

[a] TBSRC for F = 0.5 and TBSRC2 for F = 1.0. The other columns do not have a name but may be considered to be nonstandard variants of TBSRC.

[b] When r_0 = 0.5 Å and F = 0.3, optimization of **CO_3** (unprotonated form) is not converged, so it has been excluded from calculations in that case.

Table 3.6 Mean signed error (MSE) and mean unsigned error (MUE) of the QM/MM bond lengths (Å) and deprotonation energies (kcal/mol) using the TBSRC3 and TBSRCD schemes with various smearing widths r_0 (Å).

Charge scheme	r_0 (Å)	bond length (Å)		deprotonation energy (kcal/mol)	
		MSE	MUE	MSE	MUE
TBSRC3	0.5	–0.004	0.014	–0.60	2.55
	1.0	–0.004	0.014	–0.71	2.78
	2.0	–0.004	0.014	–1.10	3.50
	3.0	–0.003	0.014	–1.59	4.42
TBSRCD	0.5[a]	–0.042	0.086	1.84	2.85
	1.0	–0.012	0.030	1.58	2.14
	2.0	–0.004	0.014	0.45	1.76
	3.0	–0.005	0.015	–0.61	3.22

[a] When r_0 = 0.5 Å in the TBSBCD scheme, optimization of **CO_3** (unprotonated form) is not converged, so it has been excluded from calculations.

Table 3.7 Mean signed error (MSE) and mean unsigned error (MUE) of QM/MM bond lengths (Å) and deprotonation energies (kcal/mol) using H link atoms with the BRC scheme.

r_0 (Å)	bond length (Å)		deprotonation energy (kcal/mol)	
	MSE	MUE	MSE	MUE
0.0	–0.002	0.016	8.30	8.30
1.0	0.001	0.015	7.54	7.63
2.0	0.003	0.014	6.65	6.94
3.0	0.004	0.014	5.83	6.70
5.0	0.004	0.014	4.52	7.49

Table 3.8 MSE and MUE of the QM/MM deprotonation energies and bond lengths with the def2-TZVP and MG3S basis sets using TBSRC with a smearing width of 1.0 Å and TBRC2 without smearing.

charge scheme	basis set	bond length (Å)		deprotonation energy (kcal/mol)	
		MSE	MUE	MSE	MUE
TBSRC	def2-TZVP	–0.005	0.020	–1.16	2.18
	MG3S	–0.005	0.019	–1.73	2.43
TBRC2	def2-TZVP	0.003	0.015	–1.28	2.33
	MG3S	0.002	0.014	–1.72	2.31

Table 3.9 Mean unsigned errors in proton affinities (MUEs, in kcal/mol)

method	sites	single-point energies				with optimized geometries			
link atom		H	F[g]	F	F	H	H	F	F
tuned?[a]		no	no	ESM	CPS	no	no	CPS	CPS
smeared?		no	no	no	no	no	yes	no	yes
no MM		6	5	n.a.[b]	2.3	8	n.a	3.5	n.a
unbalanced									
SEE		19	17						
Z2		18	17						
RC[c]	M2	16	16						
balanced									
BSEE		9	5	2.4	3.4	unphys.[d]	8	unphys.	uphys.
shift[c]	M2,M2D	8	5	2.7	3.5				
Amber-1[c]	nearest M2	7	4.6	1.5	1.5				
Amber-2[c]	all	5	6	2.9	3.3				
BRCD, BRSCD[c,e]	M1.5,M2	9	6	3.2	4.2	9	8	unphys.	1.8
BRC3, BRSC3[c,e]	M2,M3	6	5	2.3	2.7	7	6	2.4	no im.[f]
BRC2, BRSC2[c,e]	all M2	6	4.5	1.5	1.7	7	7	1.6	no im.
BRC, BRSC[c,e]	M1.5	7	4.6	1.5	2.0	8	8	unphys.	1.6

[a] ESM denotes that F is tuned using the entire system model explained in paper I; CPS denotes that F is tuned on the capped primary system, i.e., without the MM subsystem.

[b] n.a. denotes not applicable

[c]The extra column ("sites") for redistributed methods shows the sites affected by redistribution (see text).
[d]unphys. denotes that some or all geometry optimizations lead to unphysical structures and do not converge.
[e]The first name given applies when there is no charge smearing, and the second name applies when smearing is used.
[f]no im. denotes that smearing leads to no improvement, i.e., the optimum smearing width is 0.
[g]In the calculations using the untuned F link atom, we used the CRENBL ECP to substitute two electrons in the core of the F link atom. The same approach was also used in paper I in the calculations on the organic molecules (i.e., the 15 molecules also tested in the present study) that used the untuned F link atom (but the other calculations in paper I for untuned F link atoms included all electrons on F)

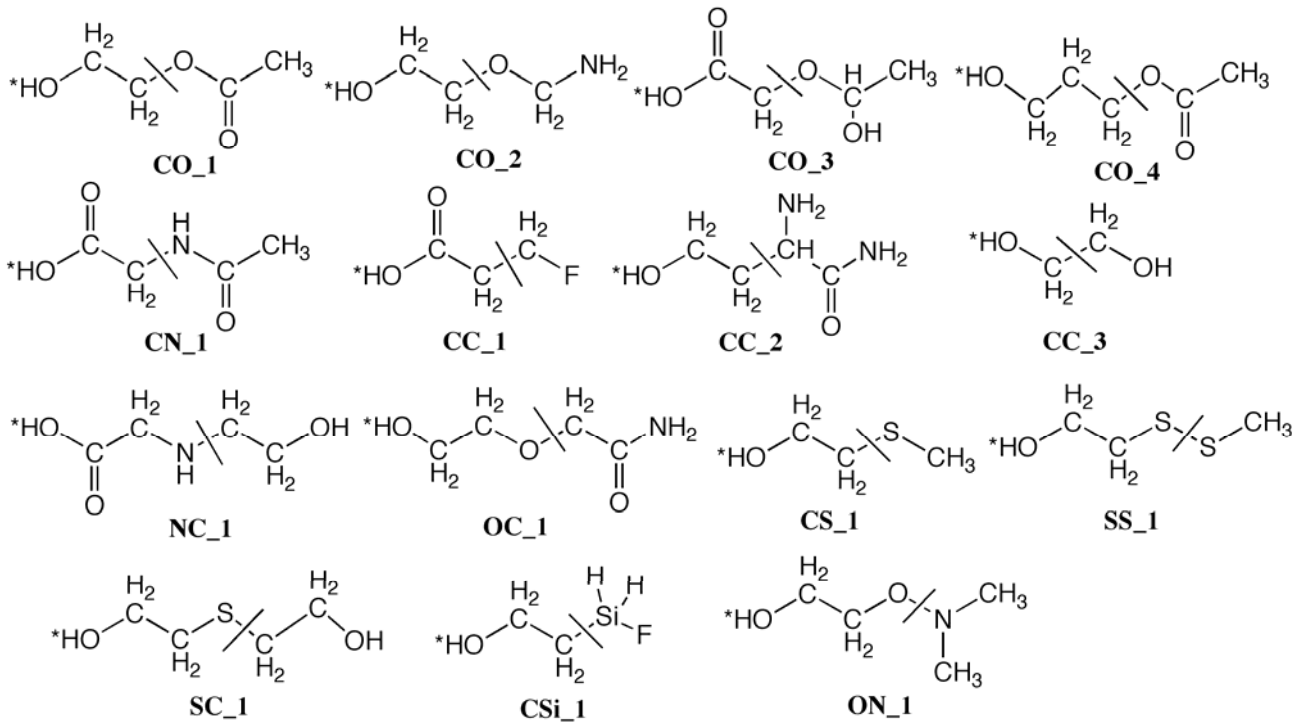


Figure 3.1 Test Suite. The asterisk * denotes the deprotonation site. The QM region is on the left of the cut bond, and the MM region is on its right.

3.8 Errata

We found a programming error in calculating the molecular mechanical energy part in the QM/MM energy calculations for the balanced charge schemes. These corrections do not change our discussion or conclusions in this chapter. The two recommended methods in the chapter, that is TBSRC with 1 Å smearing width and TBRC2 without smearing were recalculated and shown in Table 3.10.

Table 3.10 Corrected mean signed error (MSE) and mean unsigned error (MUE) in kcal/mol for deprotonation energy (kcal/mol) of 15 molecules using TBSRC with 1 Å smearing width and TBRC2

scheme	MSE	MUE
TBSRC	0.71	1.64
TBRC2	0.38	1.74

Chapter 4. Tuned and balanced redistributed charge scheme for combined quantum mechanical and molecular mechanical (QM/MM) methods and fragment methods: Tuning based on the CM5 charge model[3]

4.1 Introduction

The combined quantum mechanical and molecular mechanical (QM/MM) method is one of the most powerful methods available for simulations of large and complex systems.[1-5] One of the critical issues is the treatment of QM – MM boundaries that cut covalent bonds. Various methods have been developed to terminate the dangling bonds in the QM region, including link atoms,[6-8] orbitals,[9-11] and other kinds of capping atoms,[12-18] sometimes called pseudoatoms.

We have recently developed tuned and balanced redistributed charge schemes to better treat the QM/MM boundary.[19, 20] Mulliken charges were used to tune fluorine-like link atoms in order to obtain the correct total charge on the uncapped portion of the QM region. However, since Mulliken charges can be problematic when large basis sets are used in calculations, only small basis sets, including 6-31G* and STO-3G basis sets, were used for this tuning, and it would be problematic to extend the tuning process based on Mulliken charges to larger basis sets.

In the present study, we applied the CM5 charge model[21] to derive the tuning parameters to be used in the QM/MM simulations; this tuning process can be applied

[3] This work was supported in part by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under Award No. DE-SC0008662.

with any basis set because CM5 charges are well behaved even for extended basis sets. CM5 charges are also used as the molecular mechanical (MM) charges in the present improved method. This provides a more consistent way to do tuned and balanced redistributed charge schemes in the QM/MM calculations.

Although the tuned capping atoms will be tested in the present article only in the context of QM/MM methods, we note that the same challenge of physically capping dangling bonds on a truncated subsystem formed by cutting one or more bonds also occurs in a wide variety of fragment methods,[22] and the tuned capping atoms tested here should be equally useful for such fragment-method applications.

4.2 Methods

4.2.1 Tuned and smeared balanced redistributed charge schemes

The treatments of the boundary charges and tuned link atoms are described in previous studies,[19,20] and we only summarize the key points here. First we define M1 atoms as MM atoms directly bonded to a QM atom (through a bond that is cut by the partition) and M2 atoms as MM atoms directly bonded to M1 atoms. The balancing consists in adjusting the MM point charges on M1 atoms to conserve the total charge of the entire QM/MM system. The balanced M1 charge q_0 is then redistributed in various possible ways to avoid unphysical interactions due to MM charges being too close to the QM subsystem; in particular we here employ two schemes that were recommended previously[20] for this redistribution.

The first redistribution scheme is the balanced RC2 (BRC2) scheme,[19] in which we move equal portions of the adjusted charge to all M2 atoms.

The second redistribution scheme is the balanced and smeared redistributed charge scheme (BSRC), based on earlier method called the RC method,[23] in which we moved equal portions of the M1 charge to the midpoints of all M1–M2 bonds. In the BRC2 and

RC schemes the MM charges under consideration are point charges. In the RC scheme, the redistributed point charges at the midpoints of M1–M2 bonds are sometimes (especially when tuned F link atoms are used to capped the QM region) so close to the QM region that one encounters problems in the geometry optimizations. To alleviate this difficulty, the second charge redistribution scheme that we have recommended[20] redistributes the adjusted point charges to smeared charge distributions $\rho(r) = q_i A \exp(-2r/r_0)$, where q_i is the redistributed charge, A is the normalization factor for the charge density,[20] r is the distance to the redistributed charge center, and r_0 is the smearing width of the charge density. This is equivalent to writing the redistributed charge as[20]

$$q_i^* = q_i - q_i(1 + \frac{r}{r_0})\exp(-2r/r_0) \tag{4.1}$$

The next consideration is adding a capping atom (sometimes called a link atom or a pseudoatom) on the dangling bonds of the cut QM system. In the tuned methods, this is done with a tuned F atom that has an adjustable pseudopotential centered at its nucleus. The adjustable pseudopotential is given by[19,20]

$$U(r) = C\exp[-(r/a_0)^2] \tag{4.2}$$

where a_0 is the Bohr radius, and C is a tuning parameter adjusted to reproduce the sum of the partial charges of the uncapped portion of the QM subsystem. In previous studies,[19, 20] the partial charges were computed using Mulliken analysis with a 6-31G* basis set when the M1 atom is from the second period (Li through F) and with an STO-3G basis set otherwise. In the present study, we computed the partial charges using a more physical model, namely the CM5 charge model, which is described below.

4.2.2 CM5 charge model

Our group has developed a new charge model, namely the CM5 charge model.[21] CM5 charges are obtained by mapping the charges from Hirshfeld population analysis[24] (HPA) through the following equations.

$$q_k^{\mathrm{CM5}} = q_k^{\mathrm{HPA}} + \sum_{k \neq k'} T_{kk'} B_{kk'} \tag{4.3}$$

$$B_{kk'} = \exp[-\alpha(r_{kk'} - R_{Z_k} - R_{Z_{k'}})] \tag{4.4}$$

Where q_k^{CM5} and q_k^{HPA} are the CM5 charge and Hirshfeld charge of atom k, $r_{kk'}$ is the distance between atoms k and k', $T_{kk'}$ are parameters related to atom k and k', and R_{Z_k} is the atomic radius of atom k, and it depends only on the atomic number Z_k of atom k.

CM5 charges predict more accurate dipole moments than Hirshfeld charges and Mulliken charges, they are more stable with respect to basis sets; furthermore, as compared to charges obtained by electrostatic fitting, they are more stable with respect to conformational changes, and the equations defining them are not ill-conditioned for the charges on buried atoms. Taking into consideration these favorable features of the CM5 charge model, we applied it in the present study to tune the parameters of the link atoms as well as to obtain the MM charges.

4.3 Computational details

All computations were carried out by using the M06-2X density functional[25] for the QM subsystem. The MMFF94 force field[26] was used for the valence terms and van der Waals terms in the force field.

Three basis sets have been tested, in particular 6-31G*,[27-29] def2-TZVP,[30] and MG3S.[31] The 6-31G* basis set is used for all calculations in sections 4.4.1, 4.4.2, 4.4.4, and 4.4.5, and the other basis sets are considered in section 4.4.3. For the MM charges, we used M06-2X/CM5[21] charges of the protonated molecules with the same basis set as used for the QM subsystem. M06-2X/6-31G*/CM4M[32] charges of the protonated molecules were also tested for comparison in section 4.4.5.

For the tuning process, we used CM5 charges, but results for link atoms tuned with Mulliken charges are shown for comparison. The tuning process is performed using the protonated molecules in sections 4.4.1, 4.4.2, 4.4.3, and 4.4.5, and using the deprotonated molecules in section 4.4.4.

The test suite, shown in Fig. 4.1, contains 15 systems and is the same as that in the previous study.[20] All results in the present study are for single-point calculations without QM/MM optimizations of the structures. The geometries for the protonated molecules are the M06-2X/6-31G* optimized structures of the whole systems, and those of the deprotonated molecules are derived by deleting the protons of the protonated molecules without reoptimization.

All QM/MM calculations were carried out using our own QMMM program,[33] which is based on *Gaussian09,*[34] and a modified TINKER program.[35] All QM calculations were carried out using *Gaussian09.*[34] The CM5 charges were calculated by *CM5PAC*.[36] M06-2X/6-31G*/CM4M charges were derived from a locally modified module[37] of *Gaussian09.*[34] (The CM5 charge model, which is parametrized for the entire periodic table, will be included in an upcoming revision of *Gaussian 09*.)

4.4 Results and discussion

4.4.1 Overall performance with 6-31G* basis set

Table 4.1 shows the mean signed error (MSE) and mean unsigned error (MUE) for QM/MM deprotonation energies compared with full QM calculations using the two of recommended charge schemes,[20] the tuned BSRC scheme (TBSRC) with a smearing width of 1.0 Å and the tuned BRC2 scheme (TBRC2) without smearing. The results employing conventional hydrogen link atoms are shown for comparison.

For the tuned fluorine link atom, two different approaches were used for tuning. The first one is to use the Mulliken charge model proposed in previous studies,[19, 20] and the other is to use the CM5 charge model proposed in the current study. Table 4.1 shows that both of the tuned methods perform much better than the conventional hydrogen link atom approach. Though the schemes tuned with CM5 charges perform a little worse than the schemes tuned with Mulliken charges, a consistent 6-31G* basis set is used for tuning with CM5 charges while two different basis sets, 6-31G* and STO-3G, are used for tuning with Mulliken charges, and this consistency is expected to be both more robust and more convenient when one uses a greater variety of basis sets and considers a wider variety of systems in actual applications. Moreover, we found that unlike some cases when using the Mulliken charge model, all tuning parameters from the CM5 model are equal to or greater than 0, meaning a repulsive potential is always added to the conventional F link atom, which is quite physical since F is the most electronegative atom in the periodic table. Therefore, we believe that tuning with CM5 charges is more realistic.

Tables 4.2 and 4.3 compare the dipole moments computed from the CM5 charge model, Mulliken charge model to those calculated from the charge density for the full systems (Table 4.2) and the tuned capped primary systems (Table 4.3). Since the tuned

capped primary systems are different when tuning with Mulliken charges and with CM5 charges, we have two columns of QM dipole moments for comparison in Table 4.3. The mean unsigned errors (MUEs) of the dipole moments from the two charge models are also listed in these two tables. In order to test whether the direction, as well as the magnitude, of the dipole is correctly predicted, for each model, we also calculated the average deviation of direction (ADD) of the dipole, as defined by

$$\mathrm{ADD} = \frac{1}{14}\sum_{i=1}^{14} \left| 1 - \frac{\boldsymbol{q}_{i,\mathrm{QM}} \bullet \boldsymbol{q}_{i,\mathrm{MM}}}{|\boldsymbol{q}_{i,\mathrm{QM}}||\boldsymbol{q}_{i,\mathrm{MM}}|} \right| \tag{4.5}$$

where $\boldsymbol{q}_{i,\mathrm{QM}}, \boldsymbol{q}_{i,\mathrm{MM}}$ are the QM and MM dipole moment vectors of molecule *i*. Because of the symmetry of $HOCH_2CH_2OH$, all three components of its dipole moment are zero. Therefore, $HOCH_2CH_2OH$ is excluded from the averaging using eq 4.5, which is why only 14 molecules are used to calculate ADD. Besides the Mulliken charge model and the CM5 charge model, we also list the results computed from the MK electrostatic fitting charge model,[38] the natural population analysis (NPA),[39] and the MMFF94 force field for comparison in Table 4.2.

We found that the MUE of the dipole moments for the CM5 charge model is much smaller than that for the Mulliken charge model; that is 0.1 vs. 0.7 in the full systems and 0.2 vs. 0.5 in the tuned capped primary systems. The ADD values show that the directions of the dipole moments are also well reproduced in the CM5 charge model. This better reproduction of the fully quantum mechanical dipole moments indicates that the CM5 charge model is more physical than Mulliken analysis. As the partial atomic charges are used as an indicator for the charge distribution in the full systems and the tuned capped primary systems, CM5 charges should provide more reliable results than the previous Mulliken charges. In Table 4.2, we found that MK charges can also reproduce the QM dipole moment quite well. However, because the MK charges suffer

from the buried atom problem and are not as stable as CM5 charges, we considered the CM5 charge model as the best choice.

4.4.2 Analysis of individual molecules

To better understand the trends, individual results for all 15 deprotonation reactions are shown in Table 4.4. We show results both for employing H link atoms and for employing F link atoms tuned with CM5 charges. The QM region is in the bold characters in the table. The worst results are for the O–C and C–S bonds. However, even in these cases, the errors are smaller than the average error of using H link atoms.

To understand the trends for different QM regions and MM regions we calculated three more molecules in which a C–O bond is cut. The protonated forms of the molecules and the signed errors of the QM/MM deprotonation energies are shown in Table 4.5. Molecules CO_1, CO_4, and CO_5 contain the same kind of QM/MM boundaries and the same MM regions, but they have different QM regions. The table shows that when the QM/MM boundary is farther from the site of reaction, the error decreases significantly for both kinds of scheme (using H link atoms and using tuned F link atoms). Molecules CO_1 and CO_6 differ by one proton in the QM region, but the results for these two cases are quite different. The tuned parameters (not shown in tables) for these two cases are 0.0 vs. 0.15. Nevertheless, the schemes using tuned F link atoms perform much better than those using H link atoms in both cases. Molecules CO_1 and CO_7 contain the same QM regions and QM/MM boundaries, but have different MM regions, and their signed errors are similar.

4.4.3 Results with other basis sets

To test the sensitivity of the tuning parameters *C* and the deprotonation energies to the basis set, we performed calculations using two other basis sets, in particular MG3S and def2-TZVP.

Table 4.6 compares the tuning parameters obtained using 6-31G*, MG3S, and def2-TZVP basis sets. We found that the three sets of tuning parameters are very similar, with the typical difference being 0.05 or less. This confirms that the CM5 charge model is very stable with respect to the basis set, and we can use any basis set for tuning; this is quite different from the situation with the Mulliken charge model, in which only small balanced basis sets can be used for tuning, and one must be very careful in selecting basis sets for new elements of the periodic table. Calculations using the CM5 charge model can be reliable even if the model is tuned with a basis set different from that to be eventually used for applications. If desired, one can use a polarized double zeta basis set to do the tuning, and use these tuned parameters for simulations with larger basis sets.

Table 4.7 shows the MSE and MUE of the QM/MM deprotonation energies using MG3S and def2-TZVP basis sets, with the tuning parameters from Table 4.6. This table shows that for these basis sets, just as for the 6-31G* basis set employed in earlier sections, the schemes using tuned F link atoms perform better than those using H link atoms.

4.4.4 Tuning schemes and tuning parameters

To see how sensitive the results are with respect to the tuning scheme and tuning parameter, we also performed the tuning process using the deprotonated molecules. The procedure is similar to that for the calculations described above except that we tune to the deprotonated molecules rather than the protonated ones. Only the results for the 6-31G* basis set are shown in Table 4.8, because the other two basis sets show similar trends. Compared with the first column in Table 4.6, the tuning parameters here are smaller than those derived from the protonated molecules by around 0.3. Next, we used the averaged parameters tuned from protonated forms and deprotonated forms for calculating the deprotonation energies, the results also shown in Table 4.8. We found that the overall results are similar to those where tuning was carried out on the protonated molecules.

This indicates that it is not necessary to tune for both reactants and products since the final results are not very sensitive to the parameters. We believe that tuning on either form catches the main feature to mimic various kinds of QM/MM boundaries.

4.4.5 Various MM charges

We also carried out calculations with CM4M charges as the MM charges, shown in Table 4.9. Comparing with the results using the CM5 charges shown in Table 4.1, we found that CM4M charges give similar overall performance. This indicates that the quality of the results is not very sensitive to this kind of variation in the MM charges. But, unlike CM5 charges, CM4M charges cannot be derived for larger basis sets or for all electronic structure methods, so it is more convenient and consistent to use the CM5 charge model both to do the tuning process for the capping atoms and to derive MM charges.

4.5 Concluding remarks

In the current study, we have improved the tuned and balanced redistributed charge schemes (TBSRC and TBRC2) proposed in previous studies[19,20] as refinements of an earlier[23] redistributed charge scheme (RC) for combined quantum mechanical and molecular mechanical (QM/MM) methods. The CM5 charge model, which is not restricted to small basis sets, is applied not only to derive the MM charges used on MM atoms in the QM/MM methods but also to obtain the tuning parameters for the atoms that cap dangling bonds in the QM subsystem. This is more convenient, and it provides a more consistent way to parametrize tuned and balanced redistributed charge schemes in QM/MM calculations when using general, high-quality basis sets. We found that the CM5 charges better describe the charge distributions of molecules and hence are more physical. They are more stable with respective to basis set variations than are Mulliken

charges so that they are also more reliable and so that it is not necessary to specify the basis set when deriving the tuning parameters for the capping atom. (This makes it possible in some cases to use tuning parameters developed in previous work.) Both the TBSRC scheme and the TBRC2 scheme perform well when tuned and used with CM5 charges; in particular, comparing the results from using tuned fluorine capping atoms to those from using conventional hydrogen link atoms, we find that the tuned fluorine capping atoms reduce the errors in a test suite of deprotonation reactions by about a factor of three.

We note that tuned capping atoms may be used in other contexts as well as for QM/MM calculations. For example, they may be used for various fragment schemes,[22] such as or molecular tailoring or molecules-in-molecules methods, where one also faces the challenge of capping dangling bonds without perturbing the essential character of the electronic structure of the fragment atom. They may also be used to extend electrostatically embedded many-body methods[40] to cases where fragments are formed by cutting nondative covalent bonds.

4.6 References for Chapter 4

1. *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J.; Thompson, M. A., Eds.; ACS Symposium Series 712; American Chemical Society: Washington, DC, **1998**.
2. Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*, Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, **2000**; p 285.
3. Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
4. Senn, H. M.; Thiel, W. *Angew. Chemie Int. Ed.* **2009**, *48*, 1198.
5. Bernstein, N.; Kermode, J. R.; Csányi, G. *Rep. Prog. Phys.* **2009**, *72*, 026501.
6. Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1986**, *7*, 718.
7. Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
8. Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.

9. Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comp. Chem.* **1994**, *15*, 269.
10. Gao, J.; Amara, P.; Alhambra. C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
11. Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.
12. Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1990**, *172*, 243.
13. Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
14. Alary, F.; Poteau, R.; Heully, J. L.; Barthelat, J. C.; Daudey, J. P. *Theor. Chem. Acc.* **2000**, *104*, 174.
15. DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
16. von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.
17. Ohnishi, Y. Y.; Nakao, Y.; Sato, H.; Sakaki, S. *J. Phys. Chem. A* **2008**, *112*, 1946.
18. Lewin, J. L.; Cramer, C. J. *J. Phys. Chem. A* **2008**, *112*, 12754.
19. Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 359.
20. Wang, B.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10556.
21. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527.
22. Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2009**, *131*, 104103. Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405. Mahadevi, A. S.; Rahalkar, A. P.; Gadre, S. R.; Sastry, G. N. *J. Chem. Phys.* **2010**, *133*, 164308. Hua, S. G.; Hua, W. J.; Li, S. *J. Phys. Chem. A* **2010**, *114*, 8126. Řezač, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2010**, *6*, 91. Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336.
23. Lin, H.; Truhlar, D. G.; *J. Phys. Chem. A* **2005**, *109*, 3991.
24. Hirshfeld, F. L.; *Theor. Chem. Acc.* **1977**, *44*, 129.
25. Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157; Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
26. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
27. Hehre, W. J.; Ditchfield R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
28. Dill, J. D.; Pople, J. A. *J. Chem. Phys.* **1975**, *62*, 2921.
29. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.
30. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys*. **2005**, *7*, 3297.

31. Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, 107, 1384.
32. Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046.
33. Lin, H.; Zhang, Y.; Pezeshki, S.; Wang, B.; Truhlar, D. G. QMMM, version 1.4.0MN, University of Minnesota: Minneapolis, 2012.
34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09, version A. 02*; Gaussian, Inc.: Wallingford CT, 2009.
35. Ponder, J. W. TINKER, version 5.0, Washington University: St. Louis, 2009.
36. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *CM5PAC,* University of Minnesota: Minneapolis, MN, 2011.
37. Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM version 2011*, University of Minnesota: Minneapolis, MN, 2011.
38. Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
39. Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211. Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066. Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735. Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736.
40. E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.* **2007**, *3*, 46; E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput* **2007**, *3*, 1342; D. Hua, H. R. Leverentz,

E. A. Amin, and D. G. Truhlar, *J. Chem. Theory Comput*. **2011**, *7*, 251; E. Kurbanov, H. R. Leverentz, D. G. Truhlar, and E. A. Amin, *J. Chem. Theory Comput.* **2012**, *8*, 1.

Table 4.1 Mean signed error (MSE) and mean unsigned error (MUE) of QM/MM deprotonation energies (kcal/mol) for the test suite using H link atoms or tuned F link atoms tuned either with Mulliken charges or with CM5 charges.

scheme	H link atom		scheme	Tuned F link atom			
				Mulliken[b]		CM5	
	MSE	MUE		MSE	MUE	MSE	MUE
BSRC[a]	7.6	7.6	TBSRC[a]	1.1	1.5	1.0	2.3
BRC2	7.1	7.2	TBRC2	0.7	1.5	0.5	2.4

[a] with a smearing width of 1.0 Å. M06-2X/6-31G* is used for all calculations in this table.

[b] schemes recommended in ref. 20.

Table 4.2 Dipole moment (debye), mean unsigned error in the dipole moment (debye), and averaged deviation in direction (unitless) of the dipole of the entire system calculated from the quantum mechanical electron density and from the Mulliken, CM5, MK, NPA and MMFF94 charges.[a]

molecule	QM	Mulliken	CM5	MK	NPA	MMFF94
CO_1	3.4	3.6	3.3	3.4	4.5	4.2
CO_2	1.1	1.6	1.1	1.1	2.1	1.7
CO_3	1.5	3.1	1.7	1.5	4.0	2.3
CO_4	1.1	0.9	1.0	1.1	1.3	0.8
CN_1	0.5	1.4	0.6	0.5	1.2	1.4
CC_1	3.1	2.6	3.0	3.1	3.3	4.0
CC_2	2.4	2.5	2.6	2.4	3.4	2.3
CC_3	0.0	0.0	0.0	0.0	0.0	0.0
NC_1	5.5	5.3	5.6	5.6	6.0	5.8
OC_1	3.0	3.8	3.1	3.0	5.0	4.2
CS_1	0.4	1.8	0.4	0.4	2.7	0.5
SS_1	2.0	1.8	1.8	2.0	2.5	2.4
SC_1	1.4	4.1	1.6	1.4	5.5	1.7
CSi_1	2.6	3.9	3.2	2.6	5.4	3.6
ON_1	1.3	1.6	1.4	1.3	2.0	2.0
MUE		**0.7**	**0.1**	**0.0**	**1.3**	**0.6**
ADD		**0.28**	**0.01**	**0.01**	**0.33**	**0.04**

[a] M06-2X/6-31G* is used for all calculations in this table except for the MMFF94 column.

Table 4.3 Dipole moment (debye) and mean unsigned error in the dipole moment of the tuned capped primary system from the quantum mechanical (QM) electron density and from the Mulliken and CM5 charge models.[a]

molecule	tuned with Mulliken charges		tuned with CM5 charges	
	QM	Mulliken	QM	CM5
CO_1	2.1	2.0	1.8	1.8
CO_2	1.9	1.8	1.7	1.6
CO_3	0.7	1.4	0.6	0.6
CO_4	3.4	4.2	3.0	3.0
CN_1	5.0	5.8	4.9	4.9
CC_1	4.4	5.1	4.5	4.4
CC_2	1.7	1.9	1.7	1.5
CC_3	2.4	2.8	2.4	2.0
NC_1	5.0	5.5	4.5	4.9
OC_1	0.7	0.4	1.5	1.1
CS_1	1.6	1.7	1.6	1.5
SS_1	2.3	3.3	1.9	1.7
SC_1	1.4	2.4	1.4	1.0
CSi_1	1.6	1.8	1.5	1.4
ON_1	1.1	0.9	1.9	1.6
MUE		**0.5**		**0.2**

[a] M06-2X/6-31G* is used for all calculations in this table.

Table 4.4 QM/MM signed error (SE) and mean unsigned error (MUE) of QM/MM deprotonation energy (kcal/mol) for the test suite using the BSRC scheme with H link atoms and the TBSRC scheme with F link atoms tuned with CM5 charges.[a]

	Molecule	H link atom	Tuned F link atom
CO_1	**$HOCH_2CH_2$**-$OOCCH_3$	12.3	3.2
CO_2	**$HOCH_2CH_2$**-OCH_2NH_2	10.6	2.6
CO_3	**$HOOCCH_2$**-$OCHOHCH_3$	7.2	1.0
CO_4	**$HOCH_2CH_2CH_2$**-$OOCCH_3$	7.8	1.8
CN_1	**$HOOCCH_2$**-$NHCOCH_3$	9.3	-0.7
CC_1	**$HOOCCH_2$**-CH_2F	4.2	-1.5
CC_2	**$HOCH_2CH_2$**-$CHNH_2CONH_2$	6.5	1.5
CC_3	**$HOCH_2$**-CH_2OH	14.3	2.7
NC_1	**$HOOCCH_2NH$**-CH2CH_2OH	4.5	-1.2
OC_1	**$HOCH_2CH_2O$**-CH_2CONH_2	4.0	-5.8
CS_1	**$HOCH_2CH_2$**-SCH_3	14.0	6.9
SS_1	**$HOCH_2CH_2S$**-SCH_3	4.9	3.7
SC_1	**$HOCH_2CH_2S$**-CH_2CH_2OH	–0.1	0.7
CSi_1	**$HOCH_2CH_2$**-SiH_2F	5.4	0.6
ON_1	**$HOCH_2CH_2O$**-$N(CH_3)_2$	8.6	–0.8
MUE		**7.6**	**2.3**

[a] with a smearing width of 1.0 Å. M06-2X/6-31G* is used for all calculations in this table. The QM subsystem is bold. The QM region is in the bold characters in the table.

Table 4.5 QM/MM signed error (SE) of QM/MM deprotonation energy (kcal/mol) using the BSRC scheme with H link atoms and the TBSRC scheme with F link atoms tuned with CM5 charges.[a]

	Molecule	H link atom	Tuned F link atom
CO_5	**$HOCH_2CH_2CH_2CH_2$**-$OOCCH_3$	5.6	0.9
CO_6	**(+) $H_2OCH_2CH_2$**-$OOCCH_3$	9.8	0.6
CO_7	**$HOCH_2CH_2$**-$OOCCH_2CH_3$	12.2	3.0

[a] M06-2X/6-31G* is used for all calculations in this table. The QM region is in the bold characters in the table.

Table 4.6 Tuning parameters of the F link atom using various basis sets

	6-31G*	MG3S	def2-TZVP
CO_1	0.00	0.00	0.00
CO_2	0.20	0.20	0.20
CO_3	0.25	0.25	0.25
CO_4	0.00	0.00	0.00
CN_1	0.00	0.00	0.05
CC_1	0.70	0.70	0.70
CC_2	0.80	0.80	0.80
CC_3	0.75	0.75	0.75
NC_1	1.00	1.05	1.05
OC_1	0.80	0.90	0.85
CS_1	0.45	0.45	0.50
SS_1	0.65	0.60	0.60
SC_1	0.95	0.90	0.90
CSi_1	0.85	0.85	0.85
ON_1	0.55	0.60	0.60

Table 4.7 QM/MM mean signed error (MSE) and mean unsigned error (MUE) of QM/MM deprotonation energies (kcal/mol)[a]

Basis set	Scheme	H link atom		Tuned F link atom	
		MSE	MUE	MSE	MUE
MG3S	BSRC, TBSRC[b]	6.5	6.5	–1.6	2.6
	BRC2, TBRC2[c]	6.1	6.1	–1.9	2.5
def2–TZVP	BSRC, TBSRC[b]	6.8	6.8	–1.1	2.3
	BRC2, TBRC2[c]	6.4	6.4	–1.4	2.3

[a] The MG3S calculations were performed with the MG3S tuning parameters, and the def2–TZVP calculations were performed with the def2–TZVP tuning parameters.

[b] with a smearing width of 1.0 Å and with CM5 used for the tuning.

[c] with CM5 used for the tuning.

Table 4.8 Parameters tuned with deprotonated molecules and the signed error (SE) and mean unsigned error (MUE) for the QM/MM deprotonation energies (kcal/mol) using the averaged tuning parameters. [a]

molecule	Parameter	SE for deprotonation energy
CO_1	−0.20	2.6
CO_2	0.00	1.9
CO_3	0.05	0.7
CO_4	−0.15	1.5
CN_1	−0.30	−1.7
CC_1	0.55	−2.0
CC_2	0.45	0.3
CC_3	0.45	1.0
NC_1	0.85	−1.7
OC_1	0.75	−6.0
CS_1	0.00	5.4
SS_1	0.40	2.9
SC_1	0.80	0.3
CSi_1	0.55	−0.4
ON_1	0.45	−1.3
MUE		**2.0**

[a] M06-2X/6-31G* is used for all calculations in this table.

Table 4.9 QM/MM mean signed error (MSE) and mean unsigned error (MUE) of deprotonation energy (kcal/mol) for the test suite using BSRC and BRC2 schemes with H link atom or tuned F link atom tuned either with Mulliken charges or with CM5 charges. CM4M charges are used as MM charges.

scheme	H link atom		scheme	Tuned F link atom			
				Mulliken		CM5	
	MSE	MUE		MSE	MUE	MSE	MUE
BSRC[a]	6.7	6.8	TBSRC[a]	0.2	1.5	0.1	2.2
BRC2	6.3	6.5	TBRC2	–0.1	1.7	–0.3	2.5

[a] with a smearing width of 1.0 Å. M06-2X/6-31G* is used for all calculations in this table.

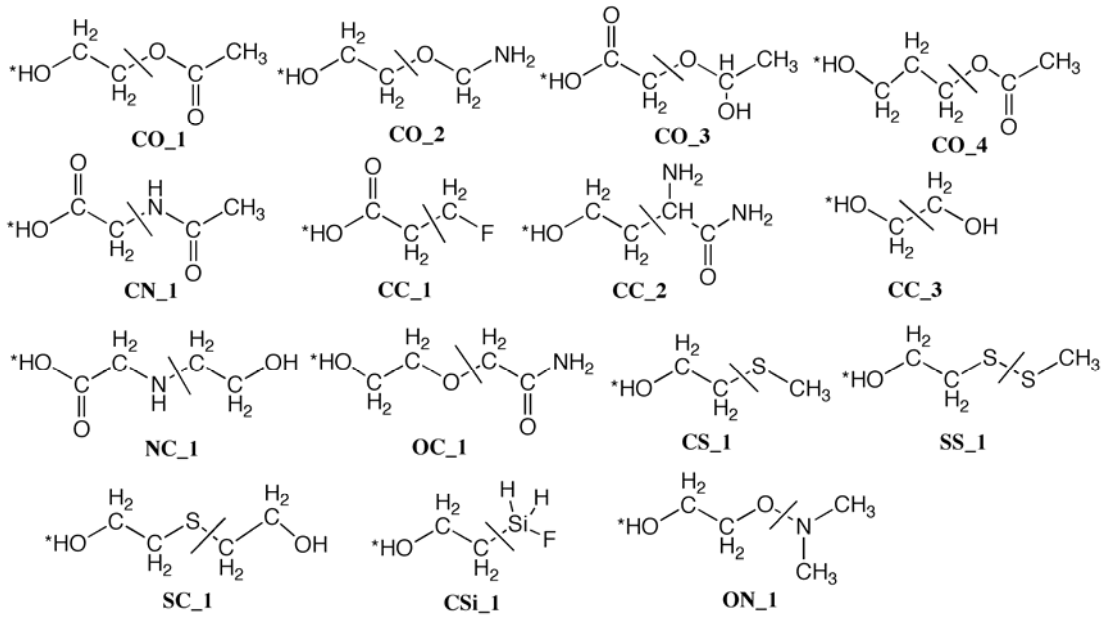


Figure 4.1 Test Suite. The asterisk * denotes the deprotonation site. The QM region is on the left of the cut bond, and the MM region is on its right. The test suite is the same as that in ref. 20.

Chapter 5. Electrostatically Embedded Molecular Tailoring Approach and Validation for Peptides[4]

5.1 Introduction

Fragment-based approaches constitute one of the most powerful classes of quantum chemistry tools for obtaining properties of a large system such as a protein. Although local density functionals (i.e., those with no Hartree-Fock exchange integrals or screened Hartree–Fock exchange and no nonlocal correlation contributions) have made it possible to compute the electronic energy of an undivided system including electron correlation with a computational cost having asymptotic scaling of N_e^3 (where N_e is the number of electrons), such calculations are still expensive to apply to large systems, and they are usually not as accurate as hybrid density functional calculations or post-Hartree-Fock correlated wave functional methods, both of which have higher scaling (N_e^4 or higher). In order to remove the N_e^3 dependency in the diagonalization procedure of the Kohn–Sham operator, Yang et al. have suggested a divide-and-conquer approach that makes it possible to avoid the N_e^3 procedure by partitioning the density matrix.[1,2,3,4,5,6,7] Other methods to reduce the scaling are based on recognizing that the most time-consuming step in the formation of the Kohn-Sham matrix is the calculation of two-electron integrals and approximating these with classical monopole-multipole or multipole-multipole interactions.[8,9,10] In the present article we consider fragmentation approaches that are equally applicable to density functional theory and all kinds of wave

[4] This work was supported in part by the National Science Foundation under grant no. CHE09-56776.

function theory, from Hartree–Fock calculations to (for example) multireference coupled cluster theory at an arbitrary excitation level.

The molecular tailoring approach (MTA)[11,12,13,14,15,0,17,0,19] is the fragment-based approach under consideration here, and we simply draw the reader's attention to other fragment methods such as the fragment molecular orbital method (FMO),[20,21,22,23,24,25,26,27,28] the molecular fractionation with conjugate caps (MFCC) method,[29,30] the effective fragment potential (EFP),[31] the systematic molecular fragmentation (SMF) method,[32,33,0] the generalized energy-based fragmentation (GEBF)[35,36,0] method, the explicit polarization (X-Pol) method,[38,39] the kernel energy method (KEM),[40, 41,42] the electrostatically embedded many-body (EE-MB) method,[43,44,45,46] the ternary interaction model,[47] the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE),[48,49] the molecules in molecules (MIM) method,[0] the multilevel fragment-based approach (MFBA),[0] the hybrid many-body interaction (HMBI) method,[52] and the many-overlapping-body expansion.53 Some of these methods, e.g., EE-MB, MIM, and GEBF, show distinctive similarities to the method proposed here and deserve special emphasis for that reason.

The MTA is applicable to wide range of systems from intermolecular interactions of small clusters to conformational stability of proteins using any electronic structure method, for example, density functional theory, post-HF methods such as CCSD(T), and semiempirical methods including the density-functional-based tight binding (DFTB) method. In the MTA, the computational time is reduced proportionally to the number of CPUs one uses, so massive parallelization may be used to greatly reduce the computational time, and this is one of the chief advantages of the MTA.

Mahadevi et al[0] have applied the MTA to benzene clusters with the MP2/6-31G++G** level of calculation for fragments, and they showed that the deviation of the electronic energy from that of full quantum mechanical (QM) calculations is within the range of chemical accuracy. Rahalkar el al.[0] applied the MTA to the alanine polypeptide, inorganic clusters including pure water, and a small protein, and they compared the performance with the FMO method. Their comparison showed that MTA provides stable accuracy with smaller dependences on the cluster or peptide size as compared with FMO method.

In order to increase the accuracy of MTA, one needs to consider several effects. The first is the combined effect of electronic polarization and long-range electrostatic interactions, which are not explicitly included in most of the previous MTA calculations, although the short-range and medium-range interactions within the fragment selected are included. For neutral species, this can lead to relatively accurate results; however, the errors may be larger in ionic or highly polar systems,[0] and it is important to consider the interfragment electrostatic interactions by including the background charges due to the other fragments. We will do that in the present paper by introducing the electrostatically embedded molecular tailoring approach (EE-MTA) in which the long-range electrostatic interactions are implicitly included based on the a procedure similar to that used in electrostatically embedded[54-56] combined QM/MM schemes.

The next consideration for possibly improving the accuracy is the type of cap atom (also called link atom) that is introduced at the fragment boundaries to satisfy the valence requirements at the bonds that are cut to make the fragments and to mimic the original electronic structure of the fragment. The most widely used cap-atom approach is

the hydrogen cap atom. One of advantages of the hydrogen cap atom is the small interaction with the rest of atoms in QM region and other fragments, and this advantage is also seen in the localized orbital approach.[57] A more advanced method than the hydrogen cap atom is the tuned atom in which a pseudo or effective core potential[58-63] is used, and parameters are adjusted to reproduce specific properties. We also mention the pseudobond method[58,59] and the quantum capping potentials method,[60] which have similar objectives. Recently, a general procedure was introduced[63] for tuning a fluorine cap atom for combined quantum mechanical and molecular mechanical (QM/MM) calculations; in this procedure a pseudopotential is added to the fluorine atom and the coefficient of the extra exponential function in the pseudopotential is optimized for each system. It was found that the tuned fluorine atom performs better than the general hydrogen cap atom, especially for the case that the partitioning into fragments occurs close to the region of the reaction center. This method could also be used in other contexts, such as the EE-MB method; here we apply it in the EE-MTA, but with a significant change. In particular, rather than tune the fluorine for each case, as in the original method, we attempt to derive general parameters suitable for use in all peptides.

The third consideration is the size of the fragments; the accuracy of the MTA or the EE-MTA should increase as one takes the sizes of fragments larger, although the quantitative effect of increasing fragment size depends on the structure of the molecule. However the computational cost becomes expensive for larger fragments, and one needs to carefully choose the size of the fragments as a compromise with the calculation cost.

The fourth consideration is the place where a molecule is partitioned; usually it is desirable to cut a nonpolar bond in which the electronegativity is similar between the two

atoms involved in the bond. In the MTA and the EE-MTA, the energies of the boundary parts of the fragments are canceled against each other, so it is expected that significant differences due to the partitioning place will not be observed. This has been shown to be true for the MTA, and here we examine it for the EE-MTA.

The improvement in going from the MTA to the EE-MTA should be important for polar and charged systems. In addition, we examine four additional considerations: (i) the size of the fragments, (ii) the place of partitioning, (iii) the effect of the type of background charge, and (iv) the type of cap atom. With regard to the cap atoms, we compare the use of widely used hydrogen cap atoms to tuned fluorine cap atoms, where the parameter of the fluorine cap atom is determined so as to minimize the error of the proton affinity as compared to full QM calculations for seven tetrapeptides. The test systems employed here for EE-MTA are the helix and sheet conformations of alanine polypeptide, Ace-$(Ala)_{20}$-NMe, where Ace is an *N*-terminal acetyl group, and NMe is a C-terminal *N*-methyl group; and the electronic energy and conformational energy are compared with those calculated by full QM calculations.

5.2 EE-MTA

5.2.1 EE-MTA with Coulomb Overcounting Term

The electronic energy of the whole peptide based on the EE-MTA is given by

$$E^{\text{EE-MTA}} = \sum E_I - \sum E_{I\cap J} + \sum E_{I\cap J\cap K} + \cdots + (-1)^{N-1} \sum E_{I\cap J\ldots\cap N} - E_{\text{OC}} \tag{5.1}$$

where E_I is the electronic energy of fragment I which is coupled with the electrostatic potential due to the rest of fragments; $E_{I\cap J}$ is the energy of the overlapping part of two

fragments, *I* and *J*; the remaining terms prior to the last term involve overlap of three or more fragments; and the last term, E_{OC}, is the sum of the Coulomb interactions that were over counted. The overcounting issue has been discussed elsewhere[33,35] and is further discussed below. In the present work we truncate eq 5.1 to second order:

$$E^{\text{EE-MTA}} \approx \sum E_I - \sum E_{I \cap J} - E_{\mathrm{OC}}. \tag{5.2}$$

This formula is best understood by an example. Consider a chain polymer with 20 monomers, which will be fragmented into 16 overlapping pentamers: 1-2-3-4-5, 2-3-4-5-6, ..., 16-17-18-19-20. There are 15 overlaps, namely 2-3-4-5, 3-4-5-6, ... ,16-17-18-19. Since the overlaps are tetramers, such a calculation is labeled as "pentamer–tetramer" fragmentation scheme, which denotes that pentamers are stitched together (by the "tailor") using tetramers. The first sum in eq 5.2 is over the pentamers, and the second sum is over the tetramers.

The over-counted electrostatics term in eq 5.2 may be calculated in more than one way. The simplest way is given by

$$E_{\mathrm{OC}} = \sum_{k}^{N_f} \sum_{l=k+M}^{N_f} \sum_{i(\in k)} \sum_{j(\in l)} \frac{Q(i,k)\ Q(j,l)}{R(i,j)} \tag{5.3}$$

where *i* and *j* are labels for atomic sites, *k* is a fragment label, $Q(i, k)$ is the partial atomic charge on atom *i* in monomer *k*, and *M* is 5 when we use the pentamer–tetramer scheme. This term is rationalized as follows: Straightforward examination of the number of occurrences of each Coulomb interaction shows that Coulomb interactions between fragments that are never contained in a common fragment are counted twice. For example, label the pentamers as 1 through 16 and label the tetramers as 2 through 16 (the label is

the number on the first fragment). Consider, as an illustration of the issue, Coulomb interactions of atoms in fragment 8 with those in fragment 11. These are counted eight times (pentamers 4-11) and subtracted seven times (tetramers 5-11), so their net count is one, which is correct. However, the Coulomb interactions between atoms in fragments 8 and 13 are counted ten times (pentamers 4–13) and subtracted only eight times (tetramers 5–8 and 10–13), so they are over counted. Therefore, we subtract all Coulomb interactions between point charges in fragment 8 and those in fragment 13. Similarly we subtract all other over-counted Coulomb interactions, as indicated in eq 5.3.

The overcounting scheme just explained will be called the straight Coulomb overcounting scheme. For our final calcuations we used a better method, but it will be easier to explain that after we give more details of the electrostatic embedding, so we postpone discussion of the final overcounting procedure to Section 5.2.2.

In the example just discussed, fourteen of the pentamers and all of the tetramers need to be capped at both ends to satisfy dangling bonds; the first and last pentamer each need only one cap each. Capping will be accomplished with cap atoms, as explained below. Each capped oligomer (i.e., pentamer or tetramer) is called a capped primary system (CPS). If, instead of pentamers, the first-order fragments of the chain polymer are trimers, we get a “trimer–dimer” calculation. If they are heptamers we get a “heptamer–hexamer” calculation, and so forth.

In the general case one might consider a tradeoff between using smaller fragments with higher orders or larger fragments with lower orders. In the present study, we did not examine higher orders of many-body interaction (higher than second order) because the tradeoff of making the fragment size smaller when one calculates higher orders of many-

body interactions would make the cap-cap atom interactions larger in the doubly capped fragments, and getting the caps too close to one another could be significant source of error or uncertainty. So it seems most reasonable to study how accurate the method can be at second order.

In the present study all fragment and dimer energies are calculated by hybrid Kohn-Sham density functional theory, in which the energy E_I of fragment I is given by

$$E_I = \frac{1}{2}\sum_{\mu}\sum_{\nu} P_{\mu\nu}\left(H_{\mu\nu}^{\text{core}} + F_{\mu\nu}\right) + \sum_{\alpha}\sum_{\beta>\alpha}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} + \sum_{\alpha,A}\frac{q_A Z_\alpha}{R_{\alpha,A}}, \tag{5.4}$$

where μ and ν represent atomic basis functions, Z_α is the charge of the nucleus of atom α of the capped primary system, CPS*, where the asterisk means the CPS is electrostatically embedded, q_A is the background charge of atom A as obtained by electrostatic potential fitting or density population analysis, $R_{\alpha\beta}$ and $R_{\alpha,A}$ is nucleus-nucleus distance of between atoms, α and β, and α and A, respectively, $P_{\mu\nu}$ is an element of the density matrix, and the core Hamiltonian is defined as

$$H_{\mu\nu}^{\text{core}} = \int \phi_\mu\left[-\frac{1}{2}\nabla^2\right]\phi_\nu d\mathbf{r} + \int \phi_\mu\left[-\left(\sum_{\alpha}\frac{Z_\alpha}{\left|\mathbf{R}_\alpha - \mathbf{r}\right|} + \sum_{A}\frac{q_A}{\left|\mathbf{R}_A - \mathbf{r}\right|}\right)\right]\phi_\nu d\mathbf{r} \tag{5.5}$$

where the first term is the kinetic energy of electrons, and the second term represents the attractive interactions of the nucleus and the electron and of the electron and the background charge. Finally $F_{\mu\nu}$ is an element of the Fock matrix, and it is given by

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda=1}\sum_{\sigma=1} P_{\mu\nu}\left[(\mu\upsilon\,|\,\lambda\sigma) - \frac{1}{2}\left(\frac{X}{100}\right)(\mu\lambda\,|\,\nu\sigma)\right] + \left(\frac{1-X}{100}\right)V_{\mu\nu}^{\text{x}} + V_{\mu\nu}^{\text{c}} \tag{5.6}$$

where X is the percentage of Hartree-Fock exchange, $(\mu\nu\,|\,\lambda\sigma)$ is a two-electron integral in

Mulliken's notation, and $V_{\mu\nu}^{\mathrm{x}}$ and $V_{\mu\nu}^{\mathrm{c}}$ are matrix elements of the local part of the exchange potential and the density-functional correlation potential, respectively.

All of the CPSs have one or two cap atoms, which may be either hydrogen atoms or tuned fluorine atoms. The latter have pseudopotentials modified from CRENBL effective core potentials.[64] These are called tuning potentials, and they have the form

$$U_{\text{tuning}}(r) = U(r) + U_0(r) + \sum_{lm}\Big[U_l(r) - U_0(r)\Big]\big|lm\big\rangle\big\langle lm\big|, \tag{5.7}$$

where $\big|lm\big\rangle\big\langle lm\big|$ are spherical harmonic projectors, U_0 and U_l are the original CRENBL potentials (representing 1s electrons) given by

$$U_l(r) = \frac{1}{r^2}\sum_j C_{lj} r^{n_{ij}} e^{-\alpha_{ij} r^2}, \tag{5.8}$$

where the parameters, C_{lj}, n_{lj}, and α_{lj} are taken from ref. 64, and $U(r)$ is an extra exponential function,

$$U(r) = C \exp[-(\frac{r}{r_0})^2] \tag{5.9}$$

where C is a coefficient to be optimized, and r_0 is taken to be 1.0 bohr. The basis set that corresponds to the core orbital of fluorine was not excluded in the previous capped atom study,[63] but here we use the effective core potential in the conventional way, that is, the basis function corresponding to the 1s orbital is replaced by the pseudopotential. As a consequence, the optimized parameter cannot be taken from the previous study.

The tuned cap F atom is placed on the bond vector of the M1 and Q1 atoms (see Fig. 5.1), as defined previously; thus it is at

$$\mathbf{r}_{\mathrm{L}} = \mathbf{r}_{\mathrm{Q1}} + d^{0}_{\mathrm{Q1-L}} \frac{\mathbf{r}_{\mathrm{M1}} - \mathbf{r}_{\mathrm{Q1}}}{\left|\mathbf{r}_{\mathrm{M1}} - \mathbf{r}_{\mathrm{Q1}}\right|} \tag{5.10}$$

where $\mathbf{r}_{\mathrm{L}}$, $\mathbf{r}_{\mathrm{Q1}}$ and $\mathbf{r}_{\mathrm{M1}}$ are the position of the cap atom, the Q1 atom, and the M1 atom, respectively; and $d^{0}_{\mathrm{Q1-L}}$ is the standard bond length as used in the previous tuned cap atom study.[64] The definitions of Q1 and M1 (and also M2 which will be used below) are illustrated in Fig. 5.1. The standard bond lengths that we employed are 1.09 and 1.01 Å for C−H and N−H bonds for using a hydrogen cap atom, and 1.33 and 1.41 Å for C−F and N−F bonds for using a tuned fluorine cap atom.

When we cap a fragment, the charge near the boundary is redistributed to avoid unrealistically large electrostatic interactions. All the tested redistribution schemes are balanced, that is, they preserve the sum of the charges of the system by requiring the sum of the embedding charges to be an integer;[73,74] this is called a balanced redistributed charge. A balanced redistributed charge scheme satisfies[73,74]

$$q^{\mathrm{ES}} = q^{\mathrm{CPS}} + q^{\mathrm{M1}} + \sum_{i \neq \mathrm{M1}} q_i^{\mathrm{SS}}, \tag{5.11}$$

where q^{ES} of the sum of charges of the entire system, q^{CPS} is the charge of the capped primary system, q^{M1} is the charge on the M1 atom, and $q_{\mathrm{i}}^{\mathrm{SS}}$ is charge of atomic site *i* in the secondary system. In a balanced calculation (all calculations reported here are balanced), q^{CPS} is zero or an integer, and the sum of the other two terms on the right hand is required to be zero or an integer. In the charge redistribution schemes used in the present study, q^{M1} is equal to zero and all embedding charges are centered at other non-CPS nuclei. Therefore, when one determines the q_i^{SS}, one needs to redistribute charges from the CPS and M1 to atomic sites of SS, excluding M1.[70,74] In the RC2 scheme,[74] the

redistributed charge is evenly redistributed over all M2 atoms, in the AMBER-2 scheme,[73] the redistributed charge is evenly redistributed over the centers of all of the secondary atoms (a secondary atom is any atom that is not a CPS atom); and in the RC3 scheme,[74] the charge is evenly redistributed on M2 and M3 atoms.

In section 5.2.2, we will need to consider combined QM-MM calculations, and as a preliminary discussion prior to that section, we need to explain the relation of electrostatically embedded fragment calculations to standard combined QM/MM calculations. In QM/MM calculations, one writes the energy as

$$E^{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{QM-MM}} + E_{\mathrm{MM-MM}} \tag{5.12}$$

where the first, second, and third terms on the right-hand side are the QM energy, the QM–MM interactions terms, and the MM–MM interaction terms. The electrostatically embedded fragment calculations discussed above are equivalent to the sum of the first two terms where only the electrostatics are included in the QM–MM term (i.e., no van der Waals or valence terms); and with the straight Coulomb overcounting scheme, we do not need the third term. In Section 5.2.2, we will, however, need the third term. In our original scheme employing redistributed charges for the QM-MM terms,[70] we used unredistributed charges in the MM–MM terms; we will call this option 1 for the MM-MM terms. In our more recent work employing balanced and redistributed charges in the QM-MM terms,[63,74] we employed balanced but unredistributed charges in the MM–MM terms (even though the charges were balanced and redistributed in the QM–MM terms), we call this option 2 for the MM–MM terms. In Section 5.2.2, we will consider a more consistent option, namely to use balanced and redistributed charges for the MM-MM term; we call this option 3 for the MM-MM terms.

One more complication must be discussed. In a typical force field, the MM–MM electrostatic interactions are not the same as the QM–MM electrostatic interactions, for example, the MM–MM electrostatic interactions do not include the 1,2- and 1,3-interactions (although they are not omitted in the QM–MM term), and (depending on the force field used) sometimes the 1,4 electrostatic interactions are excluded or reduced as well; we may call this the force field MM-MM electrostatics. However, in the discussion below, we will need to refer to calculations in which the MM–MM electrostatics include all interactions; we will denote this by referring to MM–MM electrostatics without exclusions.

5.2.2 EE-MTA with the GEBF Overcounting Correction Scheme

The overcounting scheme of eq 5.3 is called the straight Coulomb overcounting scheme. For our final calculations we used a better method, which will one called the GEBF overcounting scheme, and which is explained next.

As mentioned above, the GEBF method[35] is very similar to MTA. One difference is that GEBF includes electrostatic embedding, but employing an iterative method, in contrast to the noniterative method used here. GEBF also differs from EE-MTA in the way that fragments are assigned (it uses a distance cutoff) and in the way that one corrects from overcounting of electrostatic interactions. The overcounting difference is worth examining in more detail. In the GEBF method, eq 5.1 is replaced by

$$E^{\mathrm{GEBF}} = \sum \widetilde{E}_I - \sum \widetilde{E}_{I\cap J} + \sum \widetilde{E}_{I\cap J\cap K} + \cdots + (-1)^{N-1} \sum \widetilde{E}_{I\cap J\ldots\cap N} - \left(\sum_{k=1}^{M} C_m - 1 \right) \sum_A \sum_{B>A} \frac{Q_A Q_B}{R_{AB}} \qquad (5.13)$$

where M is the total number of constructed systems (for example, in our pentamer–

tetramer calculations, it the number of pentamers plus the number of tetramers), and C_m is the value (plus or minus one) of the coefficient in front of $\tilde{E}_{\mathrm{I}}$, $\tilde{E}_{\mathrm{I}\cap\mathrm{J}}$,…, $\tilde{E}_{\mathrm{I}\cap\mathrm{J}\ldots\cap\mathrm{N}}$ value in eq 5.13; $\tilde{E}_{\mathrm{I}}$ differs from E_I in that $\tilde{E}_I$ includes the MM–MM electrostatic interaction terms, whereas E_{I} does not, and where the final term of eq 5.13 is zero for the present application (it would not be zero for all possible fragmentation schemes, but it is zero for the present ones because $\sum_{k=1}^{M} C_m$ equals unity in the present case).

In the remainder of this section we only consider the case where $\sum_{k=1}^{M} C_m$ equals unity. In this case, the difference between eqs 1 and 13 is that the overcounting term of eq 5.3 is replaced by including the electrostatic MM-MM interaction terms (without exclusions) in the fragment energies. It is easy to see that these two schemes would be identical (i) if no covalent bonds were cut in making the fragments (thus avoiding the need for link atoms or capping atoms and meaning that the MM charges used in all steps of the calculations need neither balancing nor redistribution) or (ii) if covalent bonds are cut, but there is no balancing of MM charges and no redistribution at the boundaries. Since the overcounting term has to correct overcounting of Coulomb interactions in the QM/MM terms by means of an MM term, the correction should be as consistent as possible. Equation 5.3, however, by employing straight Coulomb interaction of unbalanced and unredistributed MM charges, is not as consistent as possible, and it would be hard to modify it to take account of adjusted charges because the adjustment on a given site depends on which fragment is being calculated.. To remedy this, we employ the GEBF overcounting scheme with the consistent option 3 mentioned at the end of the

Section 5.2.1. In this way the MM charges used to evaluate the MM-MM electrostatics are the same charges at the same locations as those used in the embedded QM calculations We tested both the straight electrostatics overcounting term and GEBF overcounting correction scheme with option 3 for the MM–MM interaction terms, and we found that the GEBF overcounting correction scheme is much more accurate. Therefore we employ this scheme in our EE-MTA method. All numerical results in this paper were obtained with the GEBF overcounting correction scheme.

5.3 Parameterization of the Pseudopotential of the Link F Atom

A tuned F atom will be denoted as F*. In the previous studies,[63] the parameter in the effective core potential of the cap F* atom was determined so as to reproduce the Mulliken charge separation estimated by a full QM calculation (the details of the procedure are given in the previous paper[63]). However, the dependence of the Mulliken charge on the choice of basis set is fairly large; in addition, the partial charges depend greatly on the choice of other methods, e.g., methods of Mulliken population analysis,[65] Hirshfeld population analysis,[66] natural bond population analysis,[67] electrostatic potential fitting,[68] or CM5,[69] that can be used to derive partial charges, and the resultant parameters depend on which method one selects. To avoid this dependence, the present parameters are determined to minimize the mean unsigned error of the proton affinity.

The advantages of the use of proton affinity for the parametrizations of the fluorine cap atom are that (i) the proton affinity, unlike the partial charge, as an observable property; therefore it is possible to use experimental data in the parametrization; (ii) one can use diffuse basis sets with a high-level of ab initio calculation which yield reasonably accurate values for the experimentally observable

proton affinity, whereas this would usually result in unphysical electron populations if one used Mulliken population analysis; and (iii) the reactant (protonated) and product (deprotonated) are well defined, and the parameter of the effective core potential depends on the both initial and final states, which makes the parametrization straightforward.

The proton affinity is calculated in the framework of the QM/MM scheme by

$$E^{\mathrm{PA,QM/MM}} = -\left[E^{\mathrm{QM/MM}}\left(\mathrm{AH}^{+}\right) - E^{\mathrm{QM/MM}}\left(\mathrm{A}\right)\right], \tag{5.14}$$

where $E^{\mathrm{QM/MM}}(\mathrm{AH}^{+})$ and $E^{\mathrm{QM/MM}}(\mathrm{A})$ represent the QM/MM energy of the protonated and deprotonated species, respectively. The QM/MM energy is defined by[70]

$$\begin{aligned} E^{\mathrm{QM/MM}} = {} & E(\mathrm{QM};\mathrm{CPS}^{**}) + \left[E(\mathrm{val};\mathrm{ES}) - E(\mathrm{val};\mathrm{CPS})\right] \\ & + E(\mathrm{Coul};\mathrm{SS}) + \left[E(\mathrm{vdW};\mathrm{ES}) - E(\mathrm{vdW};\mathrm{CPS})\right] \end{aligned} \tag{5.15}$$

where E(QM; CPS**) is quantum mechanical energy of the QM system that is coupled by the electrostatic potential to the secondary subsystem; E(val;ES) and E(val;CPS) are the MM energies for the valence interactions (for the entire system and for the CPS) corresponding to bond stretching, bond angle stretching, and torsional terms (which are described as a harmonic or trigonometric potentials in most MM force fields); E(vdW;ES) and E(vdW;CPS) are MM energies for the van der Waals interactions; and E(Coul;SS) is the Coulomb interaction energy of the secondary subsystem. Since the optimized geometry of the protonated form is used for the energy calculation of the deprotonated form, all of the MM terms are canceled in eq 5.15, and only the energy of the first term for the protonated and deprotonated form is required in the present case.

For parameterization, we used seven tetrapeptides, each of which contains two glycine residues and one other amino acid, called X; the structure is written as Ace-Gly-X-Gly-NMe, to indicate that the N terminal and C terminal are capped by Ace and NMe,

respectively. The residue X can be Arg (arginine), Ash (protonated aspartic acid), Cys (cysteine), Glh (protonated glutamic acid), Hip (protonated histidine), Lys (lysine), or Tyr (tyrosine). The hydrogen to be removed in the deprotonated form is shown in Fig. 5.2 with an asterisk. Figure 5.3 shows the partitioning scheme used for the protonation calculations.

For the parametrization, the CM5 method[69] is used to obtain background charges. The charges determined for the protonated form of the Ace-Gly-X-Gly-NMe are also used for the deprotonated form.

In total, six parameters are determined for the fluorine cap atom, depending on which kind of bond is broken and which side of the broken bond is capped. The partitioning locations are shown in Fig. 5.3. For example, part (a) shows a broken N–C_α bond capped as F*–C_α; we denote this as F*(N)–C_α. Using the same shorthand notation, the other types of capping considered are: (b) F*(C_α)–N, (c) F*(C_α)–C_O, where C_O denotes a carbonyl carbon, (d) F*(C_O)–C_α, (e) F*(C_O)–N, and (f) F*(N)–C_O. Where no confusion will result, we shorten these to F*(N)–C_α, F*–N, F*–C_O, F*–C_α, (e) F*–N, and F*–C_O. It is noted that each of these cases should have different optimized parameters. The parameters so determined are to be used for peptides with any amino acid sequence and any charge state.

All geometries are optimized for the protonated form of Ace-Gly-X-Gly-NMe by using the M06-2X density functional with the 6-31G* basis set. These geometries were started from a β sheet conformation, and hence they are not intended to represent global minima.

The CM5,[69] electrostatic potential (ESP) fitting by Merz-Kollman scheme[71] and

Mulliken[72]charges were derived using the M06-2X density functional with the 6-31G* basis set. For the calculations on the polypeptide, the CM5, ESP, and Mulliken charges are determined by full QM calculations for the whole system for each of the two conformations.

5.4 Computational details for the MTA and EE-MTA calculations

Two secondary structural motifs, namely the α helix and the β sheet, of Ace-$(Ala)_{20}$-NMe were optimized without constraints on the dihedral angles starting from the Ramachandran angles $(\varphi, \psi) = (-47^{\circ}, -57^{\circ})$ for the α helix and $(-119^{\circ}, 113^{\circ})$ for the β sheet conformation using the *AMBER99* force field,[75] where the threshold of the root mean squared (RMS) gradient was set to 0.04 kcal mol^{-1} $Å^{-1}$, and the determined geometries are shown in Fig. 5.4. These geometries are using for all of the MTA and EE-MTA calculations. All of the MTA and EE-MTA calculations were performed using the M06-2X density functional with the 6-31G* basis set.

In the present work, as mentioned in Section 5.2, we truncate the expansion at second order, as in eq 5.2. Although the inclusion of the higher-order many-body interactions should in principle provide more accurate results, the error due to the interactions of cap atoms becomes larger as the CPSs become smaller. In addition, the computational cost becomes slightly more expensive at higher order. Therefore, the second-order expansion is reasonable. The background charge used for the EE-MTA is a restrained electrostatic potential (RESP) charge[76,77] from *AMBER02*[78] which is determined by the standard AMBER procedure, in particular using B3LYP/cc-pVTZ//HF/6-31G* for a dipeptide (Ace-X-NMe, where X is each amino acid) in a continuum solvent for two important conformations, and the charge in one amino acid

unit of alanine is customized by fixing the charges of the blocking groups, as explained in ref. 78.

5.5 Results and discussion

5.5.1 Parametrization of F* against proton affinities

As the coefficient of the extra exponential in the effective core potential becomes large and positive, the electron affinity is reduced. When $C = 0$ the electron affinity of the cap atom is equal to that of fluorine atom, and when $C < 0$, the electron affinity is larger than that of the fluorine atom.

Table 5.1 gives the parameters optimized for seven peptides plus the parameters in the original CRENBL effective core potential. The optimized parameter is 0.78 for F*(M1) bound to C_α(Q1) (case (a) in Fig. 5.3), and it is 2.34 for F*(M1) bound N(Q1) (see Fig. 5.3(b)), which shows that a fluorine cap atom binding to the alpha-carbon requires a larger electron affinity as compared to that binding to the amide nitrogen to reproduce the electronic structure obtained by full QM calculations, and these *C* values both correspond to a larger electron affinity than that of hydrogen. We obtained quite different parameters, 0.78 and –3.87, for F*(M1) bound to C_α(Q1), where one of them is substituting for the amide nitrogen and the other is substituting for the C_O; the latter is electron withdrawing due to the electronegative oxygen, thus the optimized parameter is reasonable. Such a large difference also seen for the F*(M1) bound to N(Q1).

Table 5.2 gives the resultant proton affinities using the optimized parameters, and these are compared with the proton affinities calculated by using a conventional hydrogen cap atom, which is denoted as H_L. The deviations from full QM calculation are very

small with the values being under 1.0 kcal/mol for C_α-N. On the other hand, the MUE error is fairly large for N(Q1)-C_α for both tuned fluorine and hydrogen cap atom, and the large error mainly comes from positively charged Ace-Gly-Lys-Gly-NMe in the protonated form. Overall, the tuned fluorine atom shows better performance than a hydrogen atom.

5.5.2 Ace-$(Ala)_{20}$-NMe

5.5.2.1 Effect of background charge and size of fragment

Table 5.3 compares the electronic energy of the α helix to that of the β sheet, including the energy difference of these two conformations as calculated by MTA (without background charges) and by EE-MTA (with background charges). For this table the conventional hydrogen cap atom is used. In particular, Table 5.3 gives the electronic energy by MTA calculations for five sizes of fragments, and one can see that the relative energy becomes close to that of full QM when one uses larger fragments and that the energy of the α helix depends more on the size of the fragments than does the energy of the sheet. The background charges in EE-MTA remedy the high MTA relative energy of the α helix as compared with the relative energy by the full QM calculation. The effect of electronic polarization is not obvious in the β sheet conformation. Thus, the improvement of the energy difference of the two conformations comes from the improvement in energy of the α helix.

The small errors in the relative energies, when large enough fragments are considered, confirm the good performance of the molecular tailoring approach in the previous studies.[0,0] For CPS-MTA, the largest improvement of the energy difference is

seen when increasing from the trimer–dimer fragmentation (percentage error: 138%) fragmentation to the tetramer–trimer fragmentation (percentage error: 49%), and this large change is due to the α helix. This largest change can be explained by the fact that in this partition (cap-C_α…N-cap), the tetramer is the smallest fragment that that includes the important intramolecular hydrogen-bond interaction of an α helix between the *i* and *i* + 4 peptide units. Since the computational cost increases as one makes the fragment size larger, one wants to choose the smallest fragment size that gives the target accuracy, and Table 5.3 shows that pentamer–tetramer or hexamer–pentamer fragmentation is a reasonable choice for EE-MTA, although the errors with respect to the full QM calculation are only 5–6 millihartrees, respectively. The heptamer–hexamer calculation reduces this to 3 millihartrees, which is 1.9 kcal/mol or 0.1 kcal/mol per residue. We shall see further improvement in the relative energy of the two conformations in later tables as we improve the other algorithmic choices. However, already in Table 5.3, we see the encouraging result that the tetramer–trimer EE-MTA result is more accurate than the heptamer–hexamer MTA result for the difference in energy of the two structures.

5.5.2.2 Location of the cut bonds

Table 5.4 shows the electronic energy of Ace-$(Ala)_{20}$-NMe for three different partitioning positions: (i) cap-C_α(Q1)…N(Q1)-cap , (ii) cap-C_O (Q1)...C_α(Q1)-cap, and (iii) cap-N(Q1)...C_O(Q1)-cap. As can be seen in Table 5.4, the error depends on both the sizes of the fragments and the locations of the partitions. Partitions (ii) and (iii) show higher accuracy than partition (i) in only four of the ten cases. For the trimer–dimer and hexamer–pentamer cases, the partitioning between the backbone C_O and the α-carbon or between C_O and nitrogen are the best choices. Taking a broader view of the results, one

can say that the partitioning between the backbone C_O and the α-carbon is the most successful strategy. The relative accuracies of the partitions are related in part to the polarities of the cut bonds, but the common assumption that partitioning of a polar bond lowers the accuracy in only partly borne out. Since the boundary portions are canceled out in the MTA, it is expected that the large dependence on the location of the partition comes to a large extent from the electrostatic interactions of background charges, and the sensitivity of the results to overcounting correction is consistent with that.

For tetramers or larger, the least accurate location to cut a bond is usually choice (iii), which may be explained in terms of the secondary-structure hydrogen bonds of the α-helix. For schemes (i) and (ii), a tetramer contains one hydrogen bond, a pentamer contains two, etc., but for scheme (iii) a tetramer has no hydrogen bond, a pentamer has one, a hexamer has two, etc.

It is encouraging that the error in conformational energy difference is less than or equal to 6.0% in nine of 15 cases (nine of 12 cases if we exclude the trimer–dimer results); this is 1.3 millihartrees, which is less than 8 kcal/mol (0.4 kcal/mol per residue).

5.5.2.3 Charge redistribution and type of background charge

Table 5.5 shows the results of comparisons of three types of charge redistributions, which are explained at the end of Section 5.2 and which are indicated in column 2 of table 5.5. The tested charge types are: (i) RESP charge from *AMBER02*[80] (these charges differ from those of the *AMBER99* force field in which the charge is derived using Hartree-Fock level with the 6-31G* basis set and for which the dipole moment is overestimated in the gas phase[80]); (ii) CM5 charges, which are derived so as to reproduce the gas-phase dipole moments (either experimental or calculated by a high level

electronic-structure method[69]); (iii) ESP charges[71] and (iv) Mulliken charges. Note that charge types (ii), (iii), and (iv) are determined by full QM calculations for the whole system for each conformation. All four sets of charges are tabulated in supporting information, but before we discuss the differences among the results with various charges models, we should consider the charge redistribution schemes, which we do next.

As can be seen in Table 5.5, there is only a slight difference among the results for the three redistribution schemes, even for the F* cap atoms (where one might have expected a greater sensitivity due to the F* being closer to the secondary subsystem). Such a small dependency on the charge redistribution scheme has also been observed in a previous study[63] in which the various redistributed charge schemes were tested for proton affinities of variety of organic compounds.

Now we can return to the question of charge models, i.e., the different ways to obtain the charges that represents the secondary subsystems, as summarized above and as indicated in column 1 of table 5.5. These tests employ the balanced redistributed charge method, RC2. One can see in Table 5.5 that using charge distributions obtained from Mulliken analysis of the full QM calculations for each conformation yields more accurate conformational energies than does using the RESP fixed charge model for each conformation. However, using the CM5 method or ESP fitting for each conformation does not improve the results over RESP. Since the partial atomic charges of types (ii), (iii), and (iv) are determined by full quantum mechanical calculations for the whole system for each conformation, whereas in practical applications one would probably determine the charges for smaller subsystems, the results here are probably about the best we can do with fixed background charges. Further work would be required to completely

sort out the dependence on charge model, but the range of values of the errors in Table 5.5 does show that the choice of charge model can be important.

5.5.2.4 Hydrogen cap atom and tuned fluorine cap atom

Table 5.5 also compares the performance of hydrogen cap atoms to that of the tuned fluorine cap atoms for the case of pentamers stitched together by tetramers. One can see that the two kinds of cap atoms yield fairly similar deviations from a full QM calculation for the electronic energy in both conformations. One sees that the interaction between the fluorine cap atom and the PS calculated in the first term in eq 5.1 is mostly canceled in the second term. If one calculates a reaction energy such as a proton affinity, which is more sensitive to the electron density distribution, it is expected that the performance of the fluorine link atom and the H link atom will differ more, as we see in Table 5.2, where the tuned fluorine atom performs much better than the hydrogen link atom.

Eight of the ten calculations in table 5.5 lead to errors of 3.9% or less (5 millihartrees), which corresponds to 2.5 kcal/mol or 0.25 kcal/mol per residue, even though the table is restricted to fragments no larger than pentamers.

Table 5.6 shows results for the tuned F* cap atom for two locations of the cut bonds and includes hexamers and heptamers as well as pentamers, and all six results may be compared to results in Table 5.4 for the hydrogen cap atom. (A difference between hydrogen link atom and fluorine link atom is not observed for cap-Cα(Q1)…N(Q1)-cap, so the table shows results only for cap-C_O (Q1)...Cα(Q1)-cap and cap-N(Q1)...C_O(Q1)-cap.) Comparing the six cases between Tables 5.4 and 5.6, we see that the error with the fluorine cap atom is smaller for all six cases. Averaging the mean unsigned percentage

error over the six cases give an average percentage error almost three times larger (2.7% vs. 1.0%) for hydrogen cap atoms than for tuned fluorine cap atoms. This comparison shows that our attempt to parametrize a general set of tuned F* atoms for peptides was successful.

5.6 Concluding remarks

In this study, we have suggested an electrostatically embedded MTA method in which the long-range electrostatic interactions are explicitly included in an MTA calculation in the manner of electrostatically embedded QM/MM calculations with an MM overcounting term evaluated the same way as in the GEBF scheme.

We examined five variables to learn how they effect the calculated energies of two conformations (α helix and β sheet) of Ace-$(Ala)_{20}$-NMe and the calculated energy difference between these conformations: (a) polarization by extrafragment electrostatic interactions, including dependence on the type of background charge (i.e., the charge model), (b) the sizes of the fragments, (c) the locations of the sites at which the peptide is fragmented (i.e., what kind of bond is cut), (d) the redistribution of charges near the fragment boundaries, and (e) the type of the cap atom, including new generically tuned fluorine cap atoms derived in this work.

We found that electronic polarization has a large effect on the energies and relative energies; the effect is especially significant in the α helix conformation, in which the long-range interactions are very important.

We also showed the importance of taking a large enough fragment size. In the present cases, reasonably good results are obtained with tetramers, pentamers, hexamers, or heptamers as the large fragments.

The location of the fragmentation does make a difference; however, this difference is often smaller than that due to the difference associated with the sizes of the fragments.

The charge redistribution scheme has only a small effect on the results, but the choice of charge model is more significant; partial atomic charges that depend on conformation are not systematically more reliable that conformation-independent charges.

The generically tuned F* atoms perform better than hydrogen link atoms in all six cases that are sensitive to the choice of cap atoms.

Altogether we presented three tests of EE-MB for tetramer fragments, 16 tests for pentamer fragments, five tests for hexamer fragments, and five tests for heptamer fragments. Averaging over all these cases, the average error in the α–β energy difference is reduced from 49% by MTA to 17% by EE-MTA for tetramers, from 42% by MTA to 3.7% by EE-MTA for pentamer fragments, from 34% to 1.9% for hexamer fragments, and from 22% by MTA to 2.0% by EE-MTA for heptamer fragments. The average unsigned error in EE-MTA as compared to that in MTA is a factor of 2.8 lower for tetramers, a factor of 11 lower for pentamers, a factor of 18 lower for hexamers, and a factor of 11 lower for heptamers.

The new EE-MTA method can be used for any peptide, protein, or other biopolymer as well as for more general problems. One could also apply it to proteins interacting with substrates and coenzymes. Applications to a variety of problems in biological chemistry would be of great interest.

One may classify fragment methods in a variety of ways, for example, some methods are density based (they approximate the electron density by fragment methods

and then calculate an energy from the final density) and others are energy-based (they calculate the final energy from the energies of the fragments). Some methods use a single electronic structure level, and others use more than one electronic structure level[44,0,81] (e.g., different levels for different fragments or an incremental approach where the whole system is treated at a lower level and higher-order corrections are calculated based on a fragment approach) The EE-MTA method is a single-level, energy-based method, and one possible improvement is to use more than one level, as in the many-body expansion of the correlation energy.[48,49] Even among single-level energy-based methods, though, there is great variety and perhaps room for improvement. As discussed by Suarez et al.[82] and Herbert,[83] one may divide fragmentation methods into two general types: nonoverlapping and overlapping. (Mayhall and Raghavachari[53] propose a somewhat different partitioning into inclusion-exclusion methods and many-body methods.) The nonoverlapping methods are those in which each atom of the entire system (except possibly for boundary atoms in some methods) appears in at most one fragment. Single-level, energy-based nonoverlapping methods include the local self-consistent-field (LSCF) method,[84] the molecular-orbital derived empirical potential for liquids (MODEL),[85] the fragment molecular orbital (FMO) method,[20] molecular fragmentation with conjugate caps (MFCC),[29] the variational explicit polarization (X-Pol) method,[38] and many other examples. The overlapping methods are methods in which some portions of the system belong to more than one fragment. Single-level, energy-based overlapping methods include the electrostatically embedded many-body (MB) expansion,[43] the molecular tailoring approach (MTA),[13,18] the electrostatically embedded molecular tailoring approach (EE-MTA) proposed here, the closely related generalized energy-

based fragmentation (GEBF),[35,36,0,86,87] the systematic molecular fragmentation method (SMF),[88] the isodesmic fragmentation method (IFM),[89,90,91] the combined fragmentation method (CFM),[92] and others. The power of the overlapping methods is especially clear when one considers chain polymers like polypeptides and proteins, in which every residue can be strongly affected by its near neighbors on both sides along the chains—so any nonoverlapping method would necessarily treat some strong interactions as interfragment interactions, whereas an overlapping method can treat them as intrafragment interactions, which is usually more accurate. The present results for applying EE-MTA to a polypeptide are very encouraging. The overlapping methods are very general, and we see them as having enormous potential for a variety of uses. We look forward to further development, refinement, and improvements, further validation, and more applications.

5.7 References for Chapter 5

1 Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
2 Yang, W. *Phys. Rev. A* **1991**, *44*, 7823.
3 Lee, C.; Yang, W. *J. Chem. Phys.* **1992**, *96*, 2408.
4 Lu, J. P.; Yang, W. *Phys. Rev. B* **1994**, *49*, 11421.
5 York, D.; Lu, J. P.; Yang, W. *Phys. Rev. B* **1994**, *49*, 8526.
6 Zhao, Q.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 9598.
7 Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674.
8 White. C. A. ; Head-Gordon, M. *J. Chem. Phys.* **1994**, *101*, 6593
9 Strain, M. C.; Scuseria, G. E.; Frisch, N. J. *Science* **1996**, *271*, 51.
10 White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.
11 Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165.
12 Babu, K; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.
13 Ganesh, V; Dongare, R. K. ; Balanarayan, P. ;Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
14 Elango, M.; Subramanian, V.; Rahalkar, A. P.; Gadre, S. R. ; Sathyamurthy, N. *J. Phys. Chem. A* **2008**, *112*, 7699.
15 Rahalkar, A. P.; Ganesh, V. ; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101.

16 Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405.
17 Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2010**, *132*, 094102.
18 Mahadevi, A. S.; Rahalkar, A. P.; Gadre, S. R.; Sastry, G. N. *J. Chem. Phys.* **2010**, *133*, 164308.
19 Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2011**, *134*, 084111.
20 Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
21 Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
22 Kitaura, K; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* **2001**, *336*, 163.
23 Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
24 Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.
25 Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2007**, *28*, 1476.
26 Sawada, T.; Fedorov, D. G.; Kitaura, K. *Int. J. Quantum Chem.* **2009**, *109*, 2033.
27 Murata, K.; Fedorov, D. G.; Nakanishi, I.; Kitaura, K. *J. Phys. Chem. B* **2009**, *113*, 809.
28 Fedorov, D. G.; Jensen, J. H.; Deka, R. C.; Kitaura, K. *J. Phys. Chem. A* **2008**, *112*, 11808.
29 Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
30 Jiang, N; Ma, J.; Jiang, Y. *J. Chem. Phys*. **2006**, *124*, 114112.
31 Gordon, M. S.; Freitag, M. A.; Bandyopadhyay; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
32 Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
33 Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104. Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2009**, *131*, 104103.
34 Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. *Phys. Chem. Chem. Phys.*, **2012**, *14*, 7752.
35 Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.
36 Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. *J. Phys. Chem. A* **2008**, *112*, 10864.
37 Hua, S. G.; Hua, W. J.; Li, S. H. *J. Phys. Chem. A* **2010**, *114*, 8126.
38 Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108 .
39 Wang, Y.; Sosa, C.; Cembran, A.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781.
40 Huang, L.; Massa, L.; Karle, J. *J. Int. Quantum Chem.* **2005**, *103*, 808.
41 Huang, L.; Massa, L.; Karle, J. *J. Int. Quantum Chem.* **2006**, *106*, 447.
42 Huang, L.; Massa, L.; Karle, J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9541.
43 Dahlke, E. E; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
44 Dahlke, E. E; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
45 Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.
46 Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573.
47 Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.*

2005, *103*, 2255.
48 Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
49 Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.
50 Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336.
51 Řezač, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2009**, *6*, 91.
52 Wen. S.; Nanda, K.; Huang, Y.; Beran, G. *Phys. Chem. Chem. Phys*. **2012**, *14*, 7578.
53 Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669.
54 Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285.
55 Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
56 Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
57 Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
58 Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
59 Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 24114.
60 DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002,** *116*, 9578.
61 Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1990**, *172*, 243.
62 Nasuluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228.
63 Wang, B.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10556.
64 Pacios, L. F.; Christiansen, P. A. *J. Chem. Phys.* **1985**, *82*, 2664.
65 Mulliken, R. S. *J. Chem. Phys*. **1955**, *23*, 1833.
66 Löwdin, P. -O. *J. Chem. Phys.* **1950**, *18*, 365.
67 Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.
68 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., *J. Phys. Chem.* **1993**, *97*, 10269.
69 Marenich, A. V.; Jerome, S. V.; Cramer, C. J. Truhlar, D. G. *J. Chem. Theory Comput*, **2012**, *8*, 527.
70 Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.
71 Besler, B. H.; Merz Jr. K. M.; Kollman, P. A. *J. Comp. Chem.*, **1990**, *1*, 431.
72 Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
73 Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.
74 Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 359.
75 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
76 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.***1993**, *97*, 10269.
77 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.
78 Duan, Y.; Wu C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.
79 Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.
80 St.-Amant, A.; Cornell, W. D.; Kollman, P. A.; Halgren, T. A. *J. Comput. Chem.* **1995**, *16*, 1483.

81 Beran, G. *J. Chem. Phys.* **2009**, *130*, 164115.
82 Suarez, E.; Diaz, N.; Suarez, D. *J. Chem. Theory Comput.* **2009**, *5*, 1667.
83 Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
84 Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.
85 Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.
86 Hua, S.; Xu, L.; Li, W.; Li, S. *J. Phys. Chem. B* **2011**, *115*, 11462.
87 Li, W. *J. Chem. Phys.* **2013**, *138*, 14106.
88 Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
89 Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
90 Lee, A. M.; Bettens, R. P. A *J. Phys. Chem. A* **2007**, *111*, 5111.
91 Le, H.-A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2009,** *113,* 10527.
92 Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A *J. Chem. Theory Comput.* **2012**, *8*, 469.

Table 5.1 Optimized parameters of the effective core potential of a cap F* atom and parameters used in the original CRENBL effective potential.

Original CRENBL				
n_{ij} [a]	α_{ij} [a]	C_{ij} [a]	C [b]	
		U_0	(a) F*-C_α	(b) N-F*
2	2.8835	12.685306	0.78	2.34
2	3.1077	19.302589		
1	5.6122	1.002179	(c) F*-C_O	(d) C_α-F*
0	2.8146	2.245349	0.65	-3.87
		U_1	(e) F –N	(f) C_O-F*
2	44.5166	6.72324	-1.64	1.50
2	12.9487	0.929649		
1	132.4967	1.526734		

[a] taken from ref. 63.

[b] in eq 5.9.

[c] Location of the partition for (a)-(f) is shown in Fig. 5.3. C_α is an alpha-carbon, and C_O is carbonyl carbon.

Table 5.2 Signed errors of proton affinity (kcal/mol) of Ace-Gly-X-Gly-NMe calculated by EE-QM. The optimized parameters of the tuned fluorine cap atom are used. The signed errors are compared with those of EE-QM calculations with a conventional hydrogen cap atom.

	QM/MM [a,b,c]				Full QM[d]
	(a) F*-C_α		(b) N-F*		
	F*	H_L	F*	H_L	
Arg	-0.33	0.68	-5.40	-5.33	276.70
Ash	0.38	2.16	0.30	-1.03	363.40
Cys	0.53	2.84	0.20	-1.48	357.60
Glh	-0.15	1.38	1.15	-0.29	368.10
Hip	-5.04	-3.67	-1.27	-2.65	264.20
Lys	0.02	1.30	-11.69	-11.89	260.50
Tyr	-0.52	0.87	-0.05	-0.87	368.60
MUE[e]	0.99	1.84	2.87	3.36	
	(c) F*-C_O		(d) C_α-F*		
	F*	H_L	F*	H_L	
Arg	-0.19	0.27	-4.7	-4.7	
Ash	0.31	1.24	2.4	10.9	
Cys	0.38	1.48	4.0	10.0	
Glh	0.3	1.06	-1.3	5.9	
Hip	-1.64	-0.57	-5.4	4.0	
Lys	-0.13	0.43	-9.0	-12.4	
Tyr	0.22	0.96	-0.1	5.1	
MUE[e]	0.45	0.86	3.85	7.58	
	(e) F*-N		(f) C_O-F*		
	F*	H_L	F*	H_L	
Arg	-0.99	3.36	-5.82	-5.64	
Ash	0.32	7.97	5.20	4.96	
Cys	-0.04	9.54	2.91	2.92	
Glh	0.48	5.24	0.06	-0.57	
Hip	-4.92	-2.77	-0.25	-0.19	
Lys	-1.37	4.65	-14.85	-14.79	
Tyr	-0.68	3.35	1.72	1.94	
MUE[e]	1.26	5.27	4.40	4.43	

[a] The M06-2X/6-31G* basis set is used for the QM part, and the local minimum geometry is also determined by M06-2X/6-31G*.

[b] CM5 charges from full QM calculations on the protonated form (These charges are used as embedding charges for both the protonated and unprotonated molecules.)

[c] The balanced RC2 charge redistribution scheme in ref. 73 is used.

[d] M06-2X/6-31G//M6-2X/6-31G*.

[e] Mean unsigned error.

Table 5.3 MTA and EE-MTA electronic energies (in hartrees) of Ace-(Ala)20-NMe for two motifs, α helix and parallel-β sheet, and relative energy of the two motifs.[a]

Fragment unit	E_α	E_β	$E_\alpha - E_\beta$	|% error|
		CPS-MTA		
trimer (19) - dimer (18)	-5192.80567	-5192.88396	0.07829	138
tetramer (18) - trimer (17)	-5192.99501	-5192.88828	-0.10673	49
pentamer (17) - tetramer (16)	-5193.00802	-5192.88714	-0.12088	42
hexamer (16) - pentamer (15)	-5193.02488	-5192.88757	-0.13731	34
heptamer (15)- hexamer (14)	-5193.05100	-5192.88721	-0.16379	22
		EE-MTA with H_L		
trimer (19) - dimer (18)	-5193.05037	-5192.88705	-0.16332	22
tetramer (18) - trimer (17)	-5193.06990	-5192.88730	-0.18260	13
pentamer (17) - tetramer (16)	-5193.09035	-5192.88730	-0.20305	2.8
hexamer (16) - pentamer (15)	-5193.09124	-5192.88732	-0.20392	2.4
heptamer (15)- hexamer (14)	-5193.09311	-5192.88731	-0.20580	1.5
		Ref. (M06-2X/6-31G*)		
	-5193.09618	-5192.88729	-0.20889	

[a] The cap atom is H_L . The embedding charges are *AMBER* RESP charges. The charge redistrbution scheme is balanced RC2. The partitioning is cap-$C_{\alpha\ldots}$N-cap. The QM subsystems are treated by M06-2X/6-31G*.

[b] The number of fragments of a given size is in parentheses.

Table 5.4 EE-MTA electronic energies (in hartrees) of Ace-$(Ala)_{20}$-NMe for two motifs, α helix and parallel-β sheet, and relative energy of the two motifs.[a]

Fragment unit	E_α	E_β	$E_\alpha - E_\beta$	|% error|
		(i) H_L-C_α…N-H_L		
trimer (19) - dimer (18)	-5193.05037	-5192.88705	-0.16332	22
tetramer (18) - trimer (17)	-5193.06990	-5192.88730	-0.18260	13
pentamer (17) - tetramer (16)	-5193.09035	-5192.88730	-0.20305	2.8
hexamer (16) - pentamer (15)	-5193.09124	-5192.88732	-0.20392	2.4
heptamer (15)- hexamer (14)	-5193.09311	-5192.88731	-0.20580	1.5
		(ii) H_L-C_O...C_α-H_L		
trimer (19) - dimer (18)	-5193.07682	-5192.88800	-0.18881	9.6
tetramer (18) - trimer (17)	-5193.08399	-5192.88766	-0.19633	6.0
pentamer (17) - tetramer (16)	-5193.08984	-5192.88737	-0.20247	3.1
hexamer (16) - pentamer (15)	-5193.09142	-5192.88738	-0.20404	2.3
heptamer (15)- hexamer (14)	-5193.09245	-5192.88732	-0.20513	1.8
		(iii) H_L-N...C_O-H_L		
trimer (19) - dimer (18)	-5193.05835	-5192.88724	-0.17111	18
tetramer (18) - trimer (17)	-5193.02495	-5192.88540	-0.13955	33
pentamer (17) - tetramer (16)	-5193.11832	-5192.88726	-0.23106	11
hexamer (16) - pentamer (15)	-5193.09154	-5192.88703	-0.20450	2.1
heptamer (15)- hexamer (14)	-5193.08392	-5192.88734	-0.19658	5.9
		Ref. (M06-2X/6-31G*)		
	-5193.09618	-5192.88729	-0.20889	

[a] The embedding charges are *AMBER* RESP charges. The charge redistribution scheme is balanced RC2. The QM subsystems are treated by M06-2X/6-31G*.

Table 5.5 EE-MTA electronic energies (in hartrees) of Ace-$(Ala)_{20}$-NMe for two motifs, α helix and parallel-β sheet, and their relative energy.[a]

Charge	Charge RD	E_α	E_β	$E_\alpha - E_\beta$	|% error|
			H_L-C_α…N-H_L		
RESP[b]	RC2	-5193.09035	-5192.88730	-0.20305	2.8
RESP[b]	AMBER-2	-5193.09275	-5192.88986	-0.20289	2.9
RESP[b]	RC3	-5193.08949	-5192.88733	-0.20216	3.2
CM5[c]	RC3	-5193.08404	-5192.88732	-0.19672	5.8
ESP[c]	RC2	-5193.08827	-5192.88730	-0.20097	3.8
Mulliken[c]	RC2	-5193.09201	-5192.88728	-0.20473	2.0
			F*-C_α…N-F*		
RESP[b]	RC2	-5193.08859	-5192.88710	-0.20149	3.5
RESP[b]	AMBER-2	-5193.09167	-5192.88888	-0.20279	2.9
RESP[b]	RC3	-5193.08998	-5192.88702	-0.20296	2.8
CM5[c]	RC3	-5193.08162	-5192.88738	-0.19424	7.0
ESP[c]	RC2	-5193.08812	-5192.88727	-0.20084	3.9
Mulliken[c]	RC2	-5193.09196	-5192.88730	-0.20467	2.0
			Ref. (M06-2X/6-31G*)		
		-5193.09618	-5192.88729	-0.20889	

[a] The fragment size is pentamer–tetramer. The QM subsystems are treated by M06-2X/6-31G*.
[b] *AMBER* RESP charges.
[c] Determined by full QM calculation.

Table 5.6 EE-MTA electronic energies (in hartrees) of Ace-$(Ala)_{20}$-NMe for two motifs, α helix and parallel-β sheet, and their relative energy.[a]

Fragment unit	E_α	E_β	$E_\alpha - E_\beta$	|% error|
		(ii) F*-C_O...C_α-F*		
pentamer (17) - tetramer (16)	-5193.09166	-5192.88729	-0.20437	2.2
hexamer (16) - pentamer (15)	-5193.09421	-5192.88735	-0.20686	1.0
heptamer (15)- hexamer (14)	-5193.09568	-5192.88730	-0.20839	0.2
		(iii) F*-N...C_O-F*		
pentamer (17) - tetramer (16)	-5193.09468	-5192.88729	-0.20739	0.7
hexamer (16) - pentamer (15)	-5193.09283	-5192.88735	-0.20548	1.6
heptamer (15)- hexamer (14)	-5193.09544	-5192.88731	-0.20812	0.4
		Ref. (M06-2X/6-31G*)		
	-5193.09618	-5192.88729	-0.20889	

[a] The notations (ii) and (iii) in this table refer to the locations of the cut bonds, as explained near the beginning of Section 5.5.2.2; thus these results may be compared to those for the same positions of the cut bonds in Table 5.4, as discussed in Section 5.5.2.4. The embedding charges are *AMBER* RESP charges. The charge redistribution scheme is balanced RC3. The QM subsystems are treated by M06-2X/6-31G*.

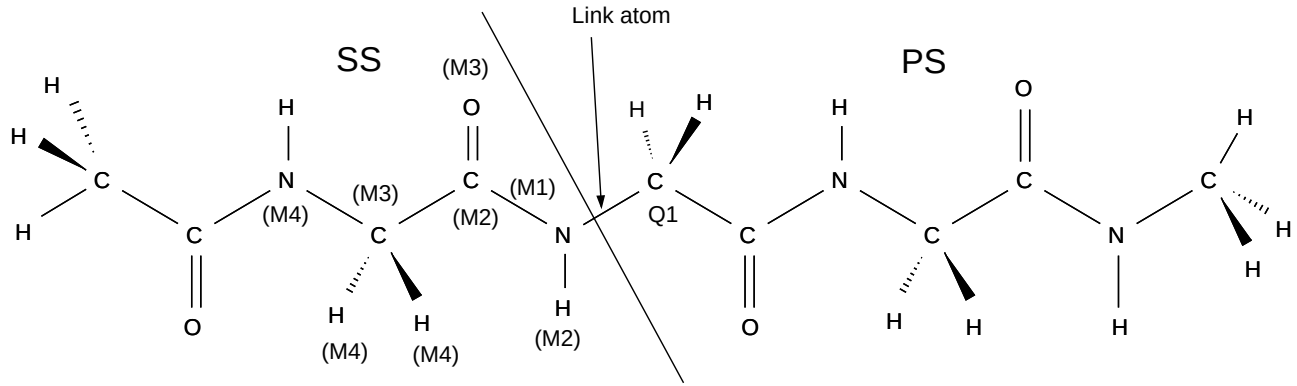


Figure 5.1 Definitions of Q1, M1, M2, M3, and M4 for the case where the interfragment bond cutting is of the cap(N)-C_O type, and the peptide is Ace-$(Gly)_3$-NMe. The cap atom is not shown, but the "link atom" arrow shows where it would be placed after the cut is made.

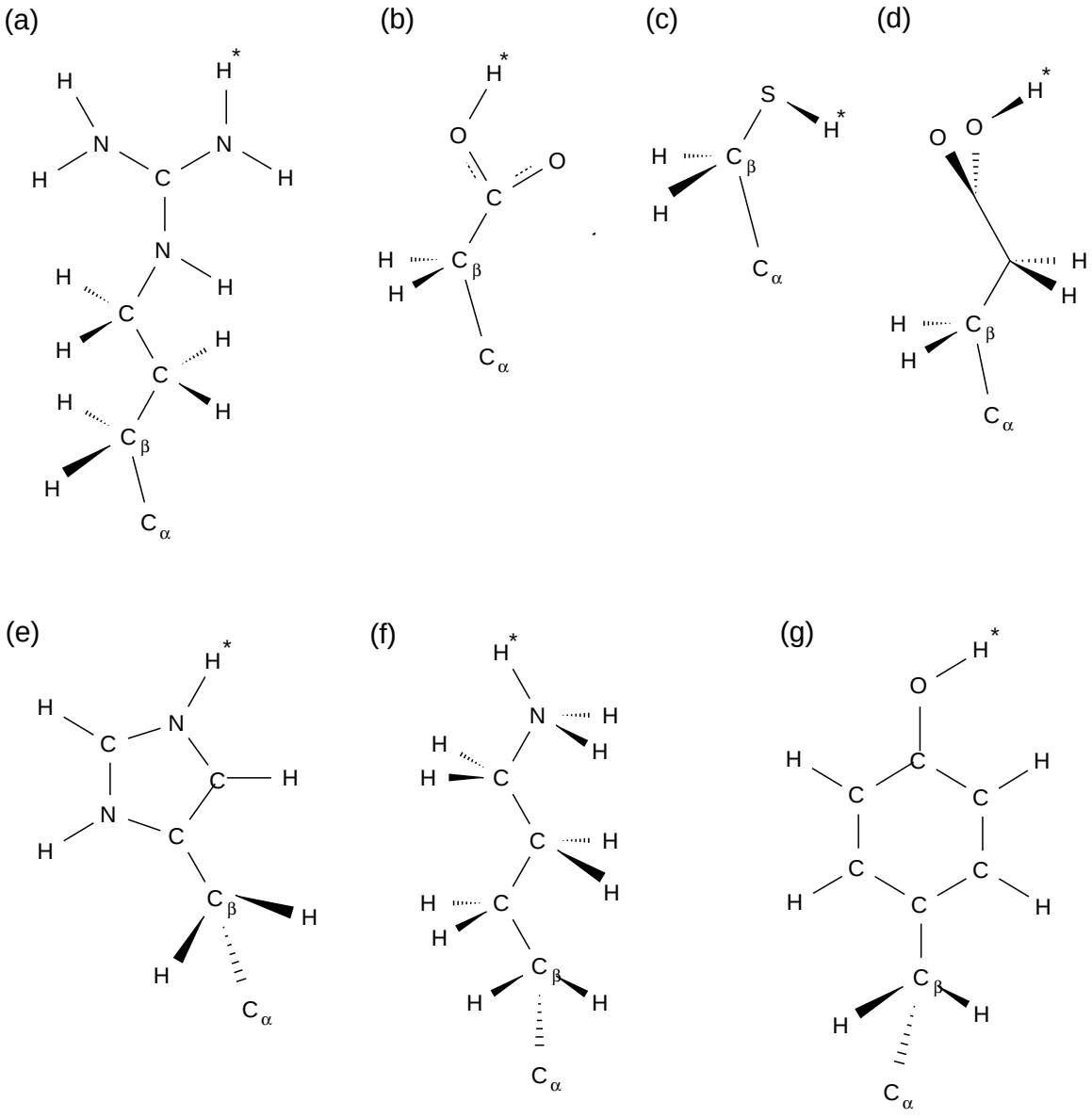


Figure 5.2 The positions of deprotonation on the side chains of X in Ace-Gly-X-Gly-NMe are given by asterisks in the illustrated side chains. (a) arginine, (b) protonated aspartic acid, (c) cysteine, (d) protonated glutamic acid, (e) protonated histidine, (d) lysine, (f) tyrosine.

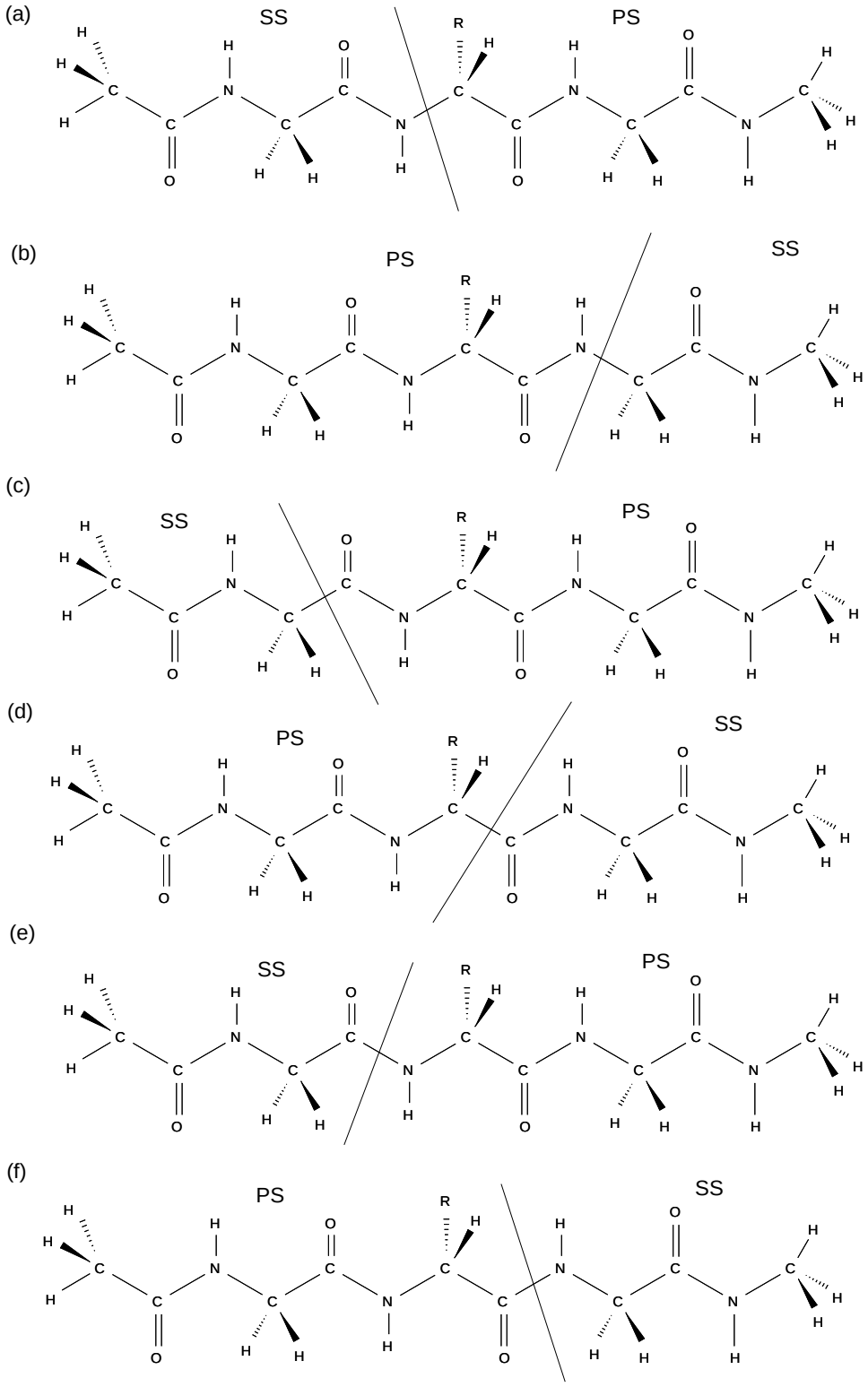


Figure 5.3 Location of the QM/MM partition in the protonation calculations. The side marked PS is the primary subsystem, which is treated by QM, and the side marked SS is the secondary subsystem, which is treated by MM.

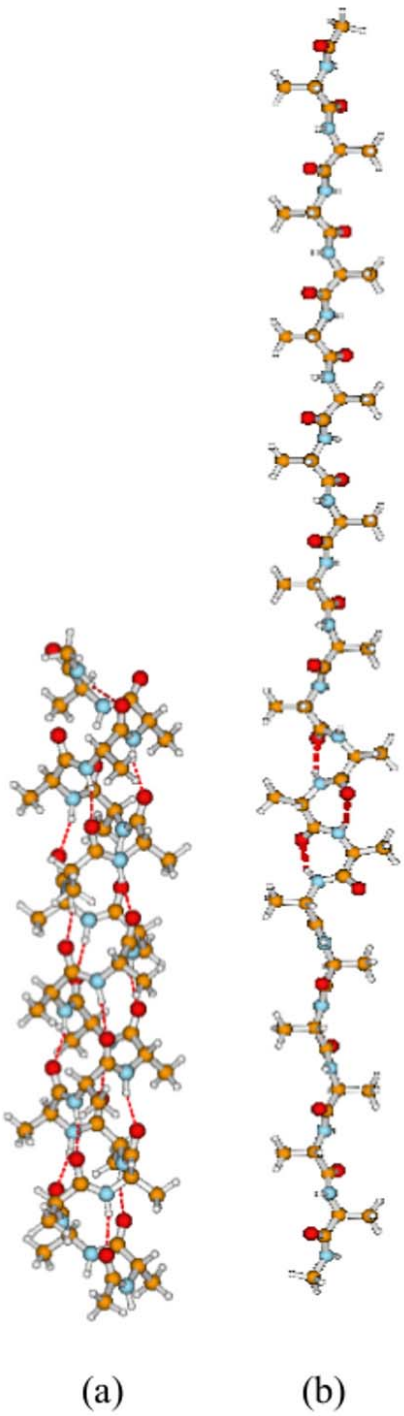


Figure 5.4 Optimized geometries of Ace-$(Ala)_{20}$-NMe for two conformations, (a) α helix and (b) parallel-β sheet, with the AMBER99 force field.

Chapter 6. Including Charge Penetration Effects in Molecular Modeling[5]

6.1 Introduction

The combined quantum mechanical and molecular mechanical (QM/MM) method is a useful tool to model large and complex systems.[1-5] It has been widely used in modeling complex molecules, condensed-phase chemistry, materials, and homogeneous, enzymatic, and heterogeneous catalysis. The electronically embedded QM/MM approach, in which interactions between QM electrons and MM partial charges are added as one-electron integrals into the QM Hamiltonian, allows the polarization of the QM electron density by the MM environment; therefore, it is a more accurate embedding scheme than mechanical embedding, in which such polarization is neglected.[6] The electrostatic interactions between the QM and MM regions are usually written as:

$$H_{\mathrm{QM/MM}}^{\mathrm{el}} = -\sum_{i,A} \frac{q_A}{r_{iA}} + \sum_{\alpha,A} \frac{Z_\alpha q_A}{R_{\alpha A}} \tag{6.1}$$

where q_A are the MM point charges; the indices i and α run over all QM electrons and nuclei, respectively; and r_{iA} and $R_{\alpha A}$ are the distances between the QM electrons and the MM point charges and the distances between the QM nuclei and the MM point charges, respectively.

The use of MM partial atomic charges as point charges at the positions of the nuclei, as in eq 6.1, is a very popular way to parameterize the electrostatics in molecular

[5] The authors thank Hannah Leverentz for providing the geometry of several dimers. This work was supported in part by the National Science Foundation under grant no. CHE09-56776 and the Air Force Office of Scientific Research under grant no. FA9550-08-1-0183.

modeling, and it is sometimes called the distributed monopole approximation. However, an MM point charge need not be a good model for the electron density of the MM subsystem. Four possible improvements can be considered, in particular the addition of higher-order multipole contributions at each nuclear center,[7-9] the use of off-nuclei point charges,[10,11] the inclusion of penetration effects,[12-14] and the treatment of additional quantum mechanical effects associated with the distributed charge distribution.[15-27] In a general way, the first three approaches all account for the same effect, namely that the actual electron density has more structure than a collection of point charges. The distributed multipole method accounts for the asymmetry of the charge distribution of an atom in a molecule, the use of off-nuclei charge centers accounts for both asymmetry and finite orbital extent, and the penetration modeling accounts for finite orbital extent. One can include both higher multipole moments and penetration effects,[20,28-30] but in the present study, we concentrate solely on the penetration effects. This has the advantage that we retain the radial, pairwise functional form for the contribution of each atom to the molecular electrostatic potential. The fourth approach is beyond the present scope, but it can include, for example, exchange repulsion or the effects of embedding atoms on the matrix elements of the QM atoms.

The essence of penetration effects, which cannot be described by MM point charges or distributed multipoles, is that when two atoms are close enough, their charge densities can overlap, and the shielding of the nuclear charge of each atom by its own electron density decreases. Various approaches have been suggested to include this effect in both MM calculations and QM/MM calculations.[13,14,16,20,28-41] In the MM studies, the MM charge densities were represented by Gaussian multipoles,[32] point charges with

damping functions,[14,20,30,41] a set of s-type Gaussian functions,[29,35,36] spherical atomic charge densities using Hartree-Fock-limit wavefunctions,[37] and single Slater-type contracted Gaussian multipole charge densities.[40] The coefficients of Gaussian type functions or damping functions were chosen to fit the electron density and electrostatic potential,[20,30,32,35-37,39,40] the electrostatic energy,[14,41] or liquid-state properties,[33] or they were based on the intermolecular overlap.[29] Similar strategies have also been used in QM/MM calculations in which charges with damping functions[14,16] and a set of s-type Gaussian functions[38] have been employed to include the charge penetration effects. Gaussian-type charge densities were used to mimic the real charge distributions of MM atoms in the link atom and double link atom methods.[13,34]

In the present work, we describe a simple scheme to include penetration effects in QM/MM calculations. Because the inclusion of distributed multipoles complicates the implementation in QM/MM calculations, and the MM point charges fitted to electrostatic-potential (ESP) can effectively include some contributions due to higher-order multipoles,[3] we do not add multipole refinements to the MM charges in this study. Moreover, it was pointed out by Cisneros et al. that damping of atom-centered point charges is more important than using distributed multipoles.[14] The basic idea of the method advanced here is using a charge screened by a Slater-type orbital[42] (STO) to represent the outermost portion of the charge density of an MM atom. The parameter used to specify the spatial extent of the screened charge is optimized to give the best agreement with Hartree-Fock electrostatic energies and with induction energies computed by symmetry-adapted perturbation theory[43] (SAPT). In section 6.2, we will derive an expression for the screened MM charge. In section 6.3, we will present the method to

optimize the exponents of the STOs for different elements, and we will present the test suite and implementation details. Section 6.4 gives the optimized parameters and an analysis of the resulting accuracy. Section 6.5 summarizes the main conclusions.

6.2 Theory

All variables and equations are in atomic units. Define a normalized STO by

$$\varphi = Ar^{n-1}\exp(-\zeta r) \tag{6.2}$$

where A is the normalization constant (normalizing φ^2 to integrate to unity), r is the distance of the electron from the nucleus, and n is the highest principal quantum number of the element. In particular, $n = 1$ for H and He, $n = 2$ for Li–Ne, $n = 3$ for Na–Ar, and $n = 4$ for K–Kr. The exponential parameter ζ is a parameter that depends on the atomic number.

We make the assumption that the charge density of an atom in the MM subsystem can be represented by two components: (i) a smeared charge Q distributed like electrons in the orbital φ and (ii) the rest of the charge, which is located at the nucleus. Since the net charge on an atom A in the MM subsystem is a parameter q_A, the charge at the nucleus is $q_A - Q$. Since the smeared charge represents the outer electrons, Q is negative and it is convenient to define $n_{\text{screen}} = -Q$. Then the charge density $\rho_A(r)$ of the smeared charge on atom A is given by

$$\rho_A(r) = -n_{\text{screen}}[Ar^{n-1}\exp(-\zeta r)]^2 \tag{6.3}$$

Figure 6.1 compares the point charge model and the screened charge model.

The exponent ζ for the STO determines the spatial extent of the smeared charge distribution of the atom. It will be optimized for several elements (H, C, N, O, F, Si, P, S,

Cl, and Br) in the present study. Our initial guess of the ζ values for these elements are the exponents of their outermost orbitals (that is, 1s for H and He, 2s for Li–Be, 2p for B–Ne, 3s for Na–Mg, 3p for Al–Ar, 4s for K–Zn, and 4p for Ga–Kr) as optimized by Clementi and Raimondi,[44] and they are shown in Table 6.1 for the elements considered in this study. Another possible choice of initial guess would be the ζ values optimized by Cusachs et al. to reproduce the overlap integrals.[45]

It can be shown that the electrostatic potential $U(r)$ at a distance *r* from the nucleus of an atom can be written as:[46]

$$U(r) = \frac{Z}{r} - 4\pi[\frac{1}{r}\int_0^r \rho(r')r'^2 dr' + \int_r^\infty \rho(r')r'\,dr'] \tag{6.4}$$

where $\rho(r)$ is the atomic electron density; and *Z* is the nuclear charge. In our screened charge model, eq 6.4 becomes

$$U(r) = \frac{q_A + n_{\text{screen}}}{r} + 4\pi[\frac{1}{r}\int_0^r \rho_A(r')r'^2 dr' + \int_r^\infty \rho_A(r')r'\,dr'] \tag{6.5}$$

where ρ_A is given by eq 6.3. (Notice that $\rho(r)$, being an electron density, is positive, where $\rho_A(r)$, being a smeared charge density is negative, just as *Q* is negative, and n_{screen} is positive.) Integrating eq 6.5 yields

$$U(r) = \frac{q_A + n_{\text{screen}} f(\zeta r)\exp(-2\zeta r)}{r} \tag{6.6}$$

where the polynomial factor *f* is

$$\begin{aligned} f(\zeta r) &= 1 + \zeta r && n = 1 \\ &= 1 + \frac{3}{2}\zeta r + (\zeta r)^2 + \frac{1}{3}(\zeta r)^3 && n = 2 \\ &= 1 + \frac{5}{3}\zeta r + \frac{4}{3}(\zeta r)^2 + \frac{2}{3}(\zeta r)^3 + \frac{2}{9}(\zeta r)^4 + \frac{2}{45}(\zeta r)^5 && n = 3 \\ &= 1 + \frac{7}{4}\zeta r + \frac{3}{2}(\zeta r)^2 + \frac{5}{6}(\zeta r)^3 + \frac{1}{3}(\zeta r)^4 + \frac{1}{10}(\zeta r)^5 + \frac{1}{45}(\zeta r)^6 + \frac{1}{315}(\zeta r)^7 && n = 4 \end{aligned} \tag{6.7}$$

The electrostatic potential of a point charge is

$$U_{\mathrm{A}}(r) = \frac{q_A}{r} \tag{6.8}$$

and the electrostatic potential of a screened charge can be written as

$$U_{\mathrm{A}}(r) = \frac{q^*_{\mathrm{A}}}{r} \tag{6.9}$$

where

$$q^*_A = q_A + n_{\mathrm{screen}} f(\zeta r)\exp(-2\zeta r) \tag{6.10}$$

We substitute q_A by q^*_A as the MM charge in eq 6.1, and the first term of eq 6.1 enters into the QM Hamiltonian in our QM/MM calculations.

Our formula is similar to eq 6.3 in ref. 14, but there are some differences. Their eq 6.3 can be rewritten as

$$q^*_A = q_A - (N_{\mathrm{val}} - q_A)\exp(\Omega_{ij}(r)) \tag{6.11}$$

where N_{val} is the number of valence electrons, and $\Omega_{ij}(r)$ is a factor depending on the distance between atoms *i* and *j*, that satisfies:

$$\begin{aligned} \Omega_{ij}(r) &= 0 \qquad (r \to 0) \\ &= -\infty \qquad (r \to \infty) \end{aligned} \tag{6.12}$$

When *r* approaches infinity, both eqs 10 and 11 become $q^*_A = q_A$; however, when *r* approaches 0, q^*_A in eq 6.10 becomes more positive, but q^*_A in eq 6.11 becomes more negative. When the MM atom becomes close to a QM nucleus, the screening effect of the electrons of the MM atoms decreases, and the QM region experiences a more positive electrostatic potential; therefore it is more physical to use eq 6.10. It seems likely that the first negative sign in eq 6.11 arose from an algebra error.

We explored two possible ways to parametrize eq 6.10. In one way, we assume that all N_{val} valence electrons are described by the STO, and we define n_{screen} as

$$n_{\text{screen}} = N_{\text{val}} - q_A \tag{6.13}$$

In the other way, since eq 6.10 will be parametrized to reproduce Hartree-Fock electrostatic and induction energies in the van der Waals region, we recognize that the parameter ζ will be suitable for describing the outermost fringe of an electron density, which may be appropriate for only one electron or a proper fraction of an electron. Moreover, the charge located at the MM nucleus equals $q_{\text{A}} + n_{\text{screen}}$, and it can be quite large if we use eq 6.13 (e.g., 6.0 for an oxygen). Since no exchange repulsion is added in the QM/MM self-consistent field (SCF) optimizations, this large charge at the MM nucleus can cause over polarization of the QM region, especially for basis sets with a large number of diffuse functions. Therefore, we reduce the number of electrons that we treat as smeared to

$$n_{\text{screen}} = \min\{N_{\text{val}} - q_A, 1\} \tag{6.14}$$

In our tests, we found that the second choice is able to represent the penetration effects while avoiding over polarization, so we only give detailed results for this choice. Looking ahead, we note that for nine of the ten elements parameterized in this article, eq 6.14 will yield $n_{\text{screen}} = 1$. The exception is hydrogen. Moreover, we found that the screening of metals (Na and Al in the test suite) does not systematically improve the results, which is possibly because the penetration effects are small for the positive charged cations. Therefore, only the nonmetals are screened in the screened charge schemes and the metals are always treated as point charges, even when they appear in the same molecule in which nonmetal charges are screened.

6.3 Methods

6.3.1 Theoretical framework

To calculate the accurate electrostatic energy and induction energy, we use symmetry-adapted perturbation theory[43] (SAPT) to partition the total interaction energy E_{int} into a sum of several physically distinct contributions. We will use three of these contributions: $E_{\text{elst}}^{(10)}$ represents the Hartree-Fock electrostatic interaction of the unperturbed monomers' charge distributions; $E_{\text{ind}}^{(20)}$ is the Hartree-Fock induction energy and equals the sum of $E_{\text{ind}}^{(20)}(A \leftarrow B)$ and $E_{\text{ind}}^{(20)}(B \leftarrow A)$, where $E_{\text{ind}}^{(20)}(A \leftarrow B)$ and $E_{\text{ind}}^{(20)}(B \leftarrow A)$ are the Hartree-Fock induction energy of monomer A with the static field of the monomer B, and vice versa; and $E_{\text{ind-exch}}^{(20)}$ is the induction exchange energy. The sum of $E_{\text{ind}}^{(20)}$ and $E_{\text{ind-exch}}^{(20)}$ is called the damped induction energy and is denoted by $E_{\text{damp-ind}}^{(20)}$ in this study.

The first question is how to compare these quantities to the components of a QM/MM calculation in which the QM method is a self-consistent-field (SCF) calculation. The SCF calculation can be either the Hartree-Fock approximation or a density functional calculation; we will use Hartree-Fock in the present article. We consider a dimer comprised of two monomers A and B. We can place either monomer A or monomer B in the QM region, and the other one is placed in the MM region. $E_{\text{elst}}^{\text{QM/MM}}$ is the interaction energy between the screened MM charges and the QM unpolarized electron density from a separate monomer calculation. $E_{\text{ind}}^{\text{QM/MM}}$ is derived in two steps. First, we place

monomer A in the QM region and monomer B in the MM region, and we compare the interaction between the screened MM charges and the QM unpolarized electron density to the interaction between the screened MM charges and the QM electron density that is relaxed in the presence of the screened MM charges. The difference is $E_{\text{ind}}^{\text{QM/MM}}(A \leftarrow B)$. Second, we switch the QM and MM regions, and calculate $E_{\text{ind}}^{\text{QM/MM}}(B \leftarrow A)$ in the same way as $E_{\text{ind}}^{\text{QM/MM}}(A \leftarrow B)$. The induction energy $E_{\text{ind}}^{\text{QM/MM}}$ is the sum of these two terms.

The comparison of $E_{\text{elst}}^{\text{QM/MM}}$ and $E_{\text{ind}}^{\text{QM/MM}}$ from QM/MM calculations to $E_{\text{elst}}^{(10)}$, $E_{\text{ind}}^{(20)}$, and $E_{\text{ind-exch}}^{(20)}$ from SAPT calculations needs special attention. The $E_{\text{elst}}^{\text{QM/MM}}$ of the QM/MM calculation can be directly compared with $E_{\text{elst}}^{(10)}$ of the SAPT calculation. However, $E_{\text{ind}}^{\text{QM/MM}}$ in QM/MM calculations doesn't have the same meaning as $E_{\text{ind}}^{(20)}$ in SAPT. In the QM/MM calculations, the induction energy $E_{\text{ind}}^{\text{QM/MM}}$ includes all orders of the perturbation of the induction, while only the second order perturbation energy is calculated in $E_{\text{ind}}^{(20)}$ in SAPT. Moreover, in SAPT calculations, the induction exchange energy $E_{\text{ind-exch}}^{(20)}$ is always evaluated and added to the induction energy $E_{\text{ind}}^{(20)}$ to give a reasonable estimation of the total interaction energy. In QM/MM calculations, it is hard to add this induction exchange energy in an empirical way, as may more readily be done for the static exchange energy, because the induction exchange energy is not even approximately pairwise additive. Therefore, we decide to compare the QM/MM induction energy $E_{\text{ind}}^{\text{QM/MM}}$ with $E_{\text{damp-ind}}^{(20)}$, that is, with the sum of the induction $E_{\text{ind}}^{(20)}$ and

induction-exchange $E^{(20)}_{\text{ind-exch}}$ energies of the SAPT calculations. The same strategy has sometimes been adopted in the development of polarizable MM force fields.[36,47]

6.3.2 Basis sets and MM charges

Two Gaussian-type basis sets were used for the QM/MM and SAPT calculations, namely the aug-cc-pVTZ basis set of Dunning and co-workers[48,49] and the def2-TZVP basis set from TURBOMOLE.[50] Although in applications one might use a smaller number of diffuse functions than are present in the aug-cc-pVTZ basis set, it is important to use a large diffuse space during parameterization to be sure that the parameterized model is stable against over polarization catastrophes.

The MM partial atomic charges are Hartree-Fock ChElPG charges[51] for the separated monomers using the same basis set as the QM method for each of the QM/MM calculations. For example, when the aug-cc-pVTZ basis set is used as the QM method in the QM/MM calculations, the MM charges are also derived using the aug-cc-pVTZ basis set.

6.3.3 Optimization methods and software

We optimize the ζ values for the ten elements (H, C, N, O, F, Si, P, S, Cl, and Br) in our test suite in order to fit the QM/MM electrostatic energies to SAPT results. The error function is based on the difference between the QM/MM and SAPT Hartree-Fock electrostatic and damped induction energies

$$\text{MUE} = \sum_{B=1}^{2} \sum_{M=1}^{\text{molecules}} \sum_{G=1}^{3} \Big(\sum_{i=1}^{2} \Big| E^{\text{QM/MM}}_{\text{elst}}(\text{QM/MM};B,M,\text{G},i) - E^{(10)}_{\text{elst}}(\text{SAPT};B,M,G) \Big| + \Big| E^{\text{QM/MM}}_{\text{ind}}(\text{QM/MM};B,M,G) - E^{(20)}_{\text{damp-ind}}(\text{SAPT};B,M,G) \Big| \Big) \tag{6.15}$$

where *B* labels the basis sets, *M* labels the molecules in the database (see below), *G* labels the geometries for each molecule in the database (see below), and *i* denotes which monomer is treated as QM. Equation 15 is minimized in the parametrization.

The SAPT calculations are performed with SAPT2008 program package[52] interfaced to the *Gaussian 03*[53] integral and self-consistent-field package. All QM/MM calculations are carried out using our own QMMM program, which is based on the *Gaussian 03*[53] and TINKER[54] programs. We use a modified version of 1.3.5.[55]

6.3.4 Test suite

We included 40 dimers in our database, as shown in Figures 6.2–6.5. Because some molecules are too large for SAPT analysis using aug-cc-pVTZ, we do calculations with the aug-cc-pVTZ basis set for only 29 out of 40 molecules in the test suite; these 29 consist of those in Figures 6.3 and 6.5 plus the ones that are labeled by stars (*) in Figure 6.2. Thus the first two sums in eq 6.15 encompass 69 cases—not 80.

We considered three geometries for each of the dimers, in particular the equilibrium geometry, a compressed geometry, and an extended geometry. The equilibrium geometry of all molecules is from several sources: i) the HB6/04, CT7/04, and DI6/04 databases[56] are the first choice, when the dimer is present in one of them. ii) $H_2O\ldots OH^-$ is the QCISD/MG3S optimized structure.[57] iii) $NH_3\ldots$HCl II, $NH_3\ldots$ClF II, and $NH_3\ldots$HF II are based on the geometry of $NH_3\ldots$ClF I in CT7/04 database[56]. For the HCl, ClF, and HF monomers in the $NH_3\ldots$HCl II, $NH_3\ldots$ClF II, and $NH_3\ldots$HF II dimers, the Cl, F, and H are placed at the same position as the Cl in $NH_3\ldots$ClF I, and the Cl–H, F–Cl, and H–F are placed along the same direction as Cl–F in $NH_3\ldots$ClF I with

the bond lengths equal 1.282 Å, 1.701 Å, and 0.924 Å, respectively; iv) Other dimers are M06-2X[58] optimized structures; the basis set we used for these optimizations is MG3S[59] for H through Cl, and 6-311+G(3d2f)[60] for Br.

In the compressed geometry, we move the monomers closer, without changing their internal structures, along a line connecting their centers of mass until the distance between the centers of mass of the two monomers is 10% *shorter* than that in the equilibrium geometry. In the extended geometry, we move the monomers farther apart along a line connecting their centers of mass until the distance between the centers of mass of the two monomers is 10% *longer* than that in the equilibrium geometry.

6.3.5 Implementation in *Gaussian 03*

As is shown in eq 6.10, the screened MM charge contains a Slater-type function. To facilitate the implementation of the Slater-type function into the Hamiltonian of the QM/MM calculations, we expand a single Slater-type function in terms of three Gaussian-type functions, as follows:[61,62]

$$\exp(-\lambda r) = \sum_{i=1}^{3} C_i \exp(-\alpha_i \lambda^2 r^2) \tag{6.16}$$

The contraction coefficients C_i and exponents α_i in eq 6.16 are listed in Table 6.2. The electrostatic potentials generated by the screened MM charges are incorporated in *Gaussian 03*[53] or *Gaussian 09*[63] as pseudopotentials using the keyword "pseudo". In the *Gaussian* implementation, the pseudopotentials do not interact with the QM nuclei. Therefore, in the QMMM program revised for this work, we calculate the interactions between the pseudopotentials and the QM nuclei, and we add them to the energy of the system.

6.4 Results and Discussion

6.4.1 SAPT results

We computed averages over the SAPT results in order to give an indication of the typical values of the various terms. The averaged SAPT results of the electrostatic, induction, and damped induction (induction plus induction-exchange) energies with two basis sets and three geometries are shown in Table 6.3. For the aug-cc-pVTZ basis set, they are averaged over 29 dimers. For def2-TZVP basis set, they are averaged over 40 dimers. To compare the results for different basis sets, we also averaged the def2-TZVP results over only 29 dimers (the same molecules as those tested by the aug-cc-pVTZ basis set), with the results shown in Table 6.4. The results in Table 6.4 can be compared with the aug-cc-pVTZ results in Table 6.3 to see the energy differences when using different basis sets.

Table 6.3 shows that the average SAPT electrostatic, induction, and damped induction energies over all geometries and basis sets are –19, –13, and –4 kcal/mol, respectively. Comparing the results from Tables 6.3 and 6.4, we found that for the electrostatic energy, the induction energy, and the damped induction energy, the differences between two basis sets are less than 7% of the averaged values. The aug-cc-pVTZ basis set contains a larger number of diffuse functions than the def2-TZVP basis set. While these functions may be important for the dispersion interactions, the electrostatic and induction energies between the monomers can be described quite well without them. Using basis sets with smaller numbers of diffuse functions not only decreases the cost of the calculations, but also it avoids the over polarization in QM/MM calculations when large charges are placed near the boundary.

Comparing three different geometries, we found that both the electrostatic and induction energies become more negative when the two monomers are placed closer, while the exchange energy becomes more positive. The absolute value of the damped induction energy is smaller than the induction energy, and it changes more slowly with geometry.

In the following discussion, we will use the electrostatic energy and damped induction energy from SAPT calculations as benchmarks to evaluate how well different charge schemes treat charge penetration effects in the QM/MM calculations.

6.4.2 QM/MM results with point charge scheme and screened charge scheme

We define the number of electrons in the STO using eq 6.14. The zeta value for each element is optimized to minimize the MUE in eq 6.15. The optimized values are shown in Table 6.5. Note that the metal elements (Na and Al) are always treated as point charges (i.e., $\zeta = \infty$). Contrary to our initial guess that the optimized zeta values would be close to the Clementi–Raimondi exponents for the outermost orbitals, they are instead similar (except for H) to half of the Strand and Bonham[64] exponents for the outermost component ($^{a}\lambda_1$ for H–Ar and $^{b}\lambda_3$ for K–Kr in their paper) of the density (the factor of 1/2 results from the fact that the screened charge used in this study has the exponent of 2ζ because the orbital is squared, not ζ); these values are also shown in Table 6.5. The mean signed error (MSE) and mean unsigned error (MUE) of the electrostatic and damped induction energies using MM point charges, MM screened charges with optimized parameters, and MM screened charges with modified Strand-Bonham (MSB)

parameters (half of the Strand-Bonham parameters for all elements except H and the optimized parameter for H) are listed in Tables 6.6 and 6.7 respectively.

Table 6.6 shows that the MUE of the electrostatic energy using the screened charge scheme with optimized parameters is 2.8 kcal/mol, compared with 8.1 kcal/mol for the traditional point charge scheme. The screened charge scheme with the MSB parameters gives an MUE of 3.1 kcal/mol, which is quite close to the optimized result. Table 6.7 shows that the MUE of the QM/MM induction energy also decreases from 1.9 kcal/mol to 1.4 kcal/mol when using the screened charge schemes, although the improvement is not as significant as the electrostatic energy (a factor of 1.4 for the induction energy vs. a factor of 2.9 for the electrostatic energy). The fact that our optimized STO exponents are very close to the values we derived from the Strand-Bonham fits to atomic densities shows that the model parameters are very physical. The results clearly show that the inclusion of penetration effects improves the description of the MM electrostatic potential and its effect on the QM system. As the penetration effects decrease exponentially as a function of the distance between atoms, the improvement is most significant for the compressed geometry, in which the penetration effects are large. For the extended geometry, in which the monomers are placed farther from each other, the difference between the point charge scheme and the screened charge scheme is relatively small.

We also tested the screened charge scheme using eq 6.13, rather than eq 6.14, to define the number of screened electrons in the Slater-type orbital, and optimized the parameters to reproduce the SAPT electrostatic energies. The electrostatic energies can be well reproduced. However, we found that the induction energy is greatly

overestimated, especially when the aug-cc-pVTZ basis set is used. The error is hard to control; therefore we abandoned eq 6.13.

One issue not so far considered is that in comprehensively parametrized force fields, one can, to some extent, make up for errors in the electrostatics by parameterization of other MM parameters or by parametrization of the MM charges. However, neither approach is satisfactory. The first approach introduces systematic errors in the QM subsystem of combined QM/MM calculations because the electrostatic terms enter the QM Hamiltonian, but the other MM terms do not affect the quantum mechanical electronic Hamiltonian. The first approach is also unsatisfactory for MM calculations because electrostatic interactions have a different functional form than the usual Lennard-Jones or Buckingham forms used for other MM nonbonded terms and therefore cannot be completely mimicked by them. The second approach introduces systematic errors because the electrostatic potential due to an unscreened Coulomb interaction with a modified charge has a different dependence on geometry than a screened Coulomb interaction with the correct charge.

6.4.3 Case studies

We use the $HCONH_2…H_2O$, $HSO_4^-…NH_4^+…H_2O$, and $HCl…H_2S$ dimers as three examples to illustrate the use of the new scheme. The equilibrium geometries of these three dimers are illustrated in Figures 6.3 and 6.4. In the calculations, either monomer 1 or 2 can be treated as the subsystem in the QM region. Tables 6.8–6.10 show the SAPT electrostatic and damped induction energies and the QM/MM electrostatic and induction energies with the three charge schemes we considered, namely, unscreened

point charges, screened charges with optimized STO parameters, and screened charges with MSB parameters for the STOs.

Table 6.8 shows the results for the $HCONH_2 \ldots H_2O$ dimer. Including penetration effects in the screened charge scheme yields a much closer match to the SAPT results. The absolute error in electrostatic energy, averaged over three geometries and two basis sets, decreases from 5.4 kcal/mol to 1.3 kcal/mol. The point charge scheme always underestimates the magnitude of the electrostatic and induction energies. Screening the point charge with the optimized parameters make the electrostatic energy more negative by 9.3 kcal/mol for the compressed geometry, 2.8 kcal/mol for the equilibrium geometry, and 0.7 kcal/mol for the extended geometry, when averaged over two QM/MM calculations with different QM regions and two basis sets.

Although the SAPT damped induction energies are similar for the aug-cc-pVTZ and def2-TZVP basis sets, the QM/MM induction energies are not. Taking the compressed geometry as an example, the SAPT damped induction energies are –4.6 and –4.5 kcal/mol for the aug-cc-pVTZ and def2-TZVP basis sets, while the QM/MM induction energies using the screened charge scheme with the optimized parameters are –6.3 and –3.7 kcal/mol for aug-cc-pVTZ and def2-TZVP basis sets. The induction energies using the aug-cc-pVTZ basis set are more negative than those using the def2-TZVP basis set in the QM/MM calculations. As the aug-cc-pVTZ basis set contains more diffuse functions than the def2-TZVP basis set, the QM region is more prone to be polarized by MM point charges when the aug-cc-pVTZ basis set is used, which leads to the increase of the induction energy. This effect is more significant for the screened charge scheme than the point charge scheme, because the charge at the MM nucleus is

larger in the screened charge scheme than that in the point charge scheme, which causes greater polarization. Despite this sensitivity, the average error in the induction energies is reduced by a about a factor of two when screening is included.

We also discuss a dimer composed of HSO_4^- and NH_4^+…H_2O, with the results shown in Table 6.9. Only the def2-TZVP basis set is used. This is a very challenging test for the charge scheme because both of the monomers are charged. From the results, we can see that the screened charge scheme improves both the electrostatic and induction energies significantly.

The results for the HCl…H_2S dimer listed in Table 6.10 are not as good as those for the two examples that we have discussed. The improvement is not very significant when H_2S is in the MM region. The performance of the screened charge scheme for individual elements will be discussed in Section 6.4.4.1.

In the examples singled out for illustration in this section, we found that the screened charge scheme with modified Strand-Bonham (MSB) parameters sometimes gives a better result than that with optimized parameters. This is because the optimized parameters are optimized for all the molecules in the test suite, rather than optimized for the specific dimers.

6.4.4 Discussion of zeta parameters

6.4.4.1 Zeta value effects on electrostatic and induction energies

In order to understand how the zeta parameters of the STOs affect the electrostatic and induction energies, we here examine the effect of varying ζ on the NH_3…NH_3 dimer. The geometry is shown in Figure 6.5. The electrostatic and induction energies are

most sensitive to the nearest MM atom, so we choose to vary the zeta parameter of the nearest MM atom. In Figure 6.6, we show the change of electrostatic energy with respect to the H zeta value with monomer 1 in the QM region and with the zeta value for N kept at its optimized value of 0.92. In Figure 6.7, we show the change of electrostatic energy with respect to the N zeta value with monomer 2 in the QM region and with the zeta value for H kept at its optimized value of 1.32. To make the comparison, we draw two lines representing the SAPT result and the QM/MM result without screening N and H respectively.

We can see that when the zeta values become larger and larger, the electrostatic energy gets closer to the result from the point charge scheme, as is easy to understand both physically and mathematically. The figures show how optimum zeta values that best reproduce the electrostatic energy are located in a reasonable range. The same trend is found for other dimers.

6.4.4.2 Consideration of individual elements

In order to understand how the screened charge scheme works for individual elements, we divided the 40 dimers into several groups. Two division schemes have been used. In the first scheme, the dimers are divided based on the elements included in the MM regions. For example, for the group corresponding to element O, we selected all the QM/MM electrostatic calculations in which one or more oxygen atom is included in the MM regions. Because an MM region always contains several different elements, each QM/MM calculation is included in several element groups. Then we calculated the MUE of electrostatic energies for different elements in their groups. The drawback of the first division scheme is that the screening effects of an MM element are not properly reflected

in the QM/MM electrostatic energy when the MM element is far from the QM region. Therefore, in the second scheme, only the MM atoms that are close to the QM region are considered. We calculated the distances between all QM and MM atom pairs, and identified the shortest QM–MM distance. We define a close MM atom as one that has a distance from any QM atom that is less than the sum of 0.30 Å and the shortest QM–MM distance. Then we divided the 40 dimers based on the elements among the close MM atoms. We carried out the division only for the equilibrium geometry, and we assigned the same group membership for the compressed and extended geometries as in the equilibrium geometry. Because the QM/MM electrostatic energy is most sensitive to the screening of the closest MM atoms, the second division scheme is more appropriate to test the contribution of individual elements. Note that if we change 0.30 Å in the criterion to define a close MM atom to infinity, the second division scheme becomes the first division scheme. We show the number of calculations included in each group and the MUE of the QM/MM electrostatic energy for both the two division schemes in Table 6.11. We define an improvement ratio (IR) as

$$\mathrm{IR} = \frac{\mathrm{MUE\ (point\ charge\ scheme)}}{\mathrm{MUE\ (screened\ charge\ scheme)}} \tag{6.17}$$

IRs for individual elements using the screened charge scheme with the optimized parameters are also shown in Table 6.11, and it is gratifying that they are all greater than or equal to 1.6. The results of the two division schemes show similar trends (except for C, for which the first division scheme does not reflect the screening effect of the C atom).

For the point charge scheme, Cl and N have the largest MUEs. All of them are greater than 10 kcal/mol for both the two division schemes. The large error of the Si

using the first division scheme is mainly due to other elements, as Si is not a close MM atom in any dimers.

For the screened charge scheme with optimized parameters, N is the only element with an MUE of more than 3 kcal/mol by either division, whereas when screening is not employed all elements have an MUE greater than or equal to 3.6 kcal/mol for both divisions. We found that the large error of N is mainly due to the inclusion of NH_3…HF, NH_3…HCl, and 2T(Al)…NH_3 dimers, in which the monomers are very close to each other.

Next, we compare the improvement ratios for individual elements. In the second division scheme, the elements that have the largest improvement ratios are C (7.7), Cl (6.9), and F (4.9); followed by H (3.8), Br (2.6), and O (2.1); N (1.9), P (1.7), S (1.7) have the smallest improvement, but even in these cases, the improvement is significant. These findings show that for all elements, electrons distributed in an STO can mimic the charge distribution of the MM atoms well enough to improve the results as compared to a point charge model.

6.4.5 An application to ten water dimers with a variety of geometries

To test the broader usefulness of the new method, we applied it to water dimers with ten geometries tested by Reinhardt et al.[65] They have carried out SAPT calculations on these dimers using the aug-cc-pVTZ basis set. We carried out QM/MM calculations with the aug-cc-pVTZ basis set on their geometries using the point charge scheme and the screened charge scheme with MSB parameters, and we compared the results with their benchmark results of electrostatic and induction energies. The point charges of the water monomer are Hartree-Fock ChElPG charges[51] calculated with the aug-cc-pVTZ

basis set. Figures 6.8 and 6.9 compare the electrostatic and induction energies for these 10 geometries. Because either monomer in the dimer can be treated as the QM region, we have two QM/MM electrostatic energies at each geometry. We found that using the screened charge scheme decreases the MUE of the electrostatic energy from 1.42 kcal/mol to 0.42 kcal/mol and the MUE of the induction energy from 0.11 kcal/mol to 0.06 kcal/mol. Note that the averaged electrostatic and damped induction energies of the ten water dimers are –5.48 kcal/mol and –0.57 kcal/mol, respectively. Therefore, the percent error drops from 26% to 8% and from 19% to 11% for the QM/MM electrostatic and induction energies respectively. These findings show that including penetration effects is important for accurate modeling. We also carried out calculations using the screened charge scheme with optimized parameters, in which the zeta value for O is 1.20 rather than 1.12 as in the MSB parameters. The MUE of electrostatic energy is 0.57 kcal/mol, which is larger than that using MSB parameters, but still much better than that for point charges, and the MUE of the induction energy is 0.05 kcal/mol. We found that if the two monomers in the dimer are not very close (two monomers are very close in the case of $H_2O\ldots OH^-$ in the test suite for parametrization), the zeta value of 1.12 for O gives better results than the optimized value from Table 6.5.

6.4.6 Limitations of the scheme

We note five limitations.

First, only a single STO with a fixed zeta value is used for each element in our scheme. However, a single STO may not describe the electron distribution well in all regions. When the QM electron goes closer to the MM atom, it should feel a charge distribution with a larger zeta value. This can be improved by double zeta STOs, in which

a more diffuse STO is used to describe the ultimate outer layer and a less diffuse one to describe the penultimate layer. Moreover, the zeta value for an element changes when the element has different partial charges and bonding environments, as pointed out before,[66] so the zeta value we optimize here is only an average value for an element in different environments. If one is willing to treat zeta as an MM parameter and optimize it separately for different hybridization states (e.g., sp^2 O would have a different value than sp^3 O) or different functional groups or charge states, one could do even better.

Second, to avoid over polarization, we use at most one electron in the STO. A small number of electrons in the STO may underestimate the charge penetration effect, especially for the compressed geometry. One may include more electrons in the STO if the over polarization of the QM region can be avoided (e.g., when def2-TZVP basis set is used for the QM calculations); this would, for example, improve the results in Table 6.10 when H_2S is the MM region and S is the close MM atom.

Third, when multiply charged cations (such as Zn^{2+} and Mg^{2+}) are placed in the MM region near the QM–MM boundary, we found that the QM region may be greatly over polarized by the MM region if a large number of diffuse functions are used for the QM region, such as that in the aug-cc-pVTZ basis set. This is due to the lack of exchange and orthogonality interactions between MM and QM regions in the QM/MM SCF optimizations. Although the problem is much less severe with a less diffuse basis set, our method cannot completely solve this problem. *Ab initio* model potentials[67,68] (AIMP), smeared charges,[13,34] and damped charges,[38] can be applied to alleviate the over polarization. One can also avoid the problem by not placing a QM–MM boundary next to a multiply charged MM cation.

Fourth, we have not tested the method for rare gases.

Fifth, we have so far developed the method only for the case where a QM/MM boundary does not pass through a covalent bond.

If desired, further studies can be carried out to remove limitations 1, 2, 4, and 5 mentioned above. Despite limitations 1–3, the method already provides a much more realistic treatment of the electrostatic and induction energies in the general case.

6.5 Conclusions

In this article, we proposed a general screening scheme to include charge penetration effects in the treatment of electrostatic interactions in molecular modeling, and we parametrized it and applied in by using electronically embedded QM/MM calculations. Our scheme utilizes a Slater–type orbital to mimic the outer portion of the electron distribution around the MM atom. By including the charge penetration effects, we can greatly improve the description of electrostatic interactions in the QM/MM method. The parameters for the STOs of several common used elements (H, C, N, O, F, Si, P, S, Cl, and Br) are optimized to reproduce the SAPT electrostatic and damped induction energies. For the metal elements, we suggest to keep the point charge scheme as no systematic improvement has been found by treating them as screened.

We found that the optimal exponential parameters are very close to the values describing the outmost layer of the electron density of atoms in Strand and Bonham's fits to atomic electron densities. This is extremely encouraging in that it shows that the method is very physical. The finding that screening does not offer systematic improvement for metal atoms is also physical since metal atoms in molecules usually have a partial positive charge and hence less diffuse electron density. These findings,

combined with the availability of Strand and Bonham's fits for the first 36 elements, means that parameters for the nonmetals (except rare gases) with $Z \leqslant 36$ that are not optimized here (B, Ge, As and Se) can be obtained from their fits to electron densities.

Since the point charge model has been used in the overwhelming majority of MM parameterizations, but the present work shows that the point charge model leads to systematic errors in electrostatics and induction energies, we conclude that conventional MM parameterizations can only succeed by systematic cancellation of errors. Improving the electrostatics by including charge penetration effects can in principle lead to a new generation of more physical MM parameter sets.[35,69] By evaluating the electrostatic energies more accurately, we can derive more accurate and physical empirical parameters for the exchange repulsion and dispersion. The present work shows how this can be done in a practical way. In particular the bare Coulomb interaction is replaced by a central potential centered at the nuclei with only one additional parameter per element.

Although the formulation of the electrostatic interactions in the current work applies only to QM/MM calculations, the formula for MM–MM electrostatic energy calculations can also be derived based on the screened charge model. One obtains

$$E = \int dr_1 \int dr_2 \frac{\rho_A(r_1)\rho_B(r_2)}{|r_1 - r_2|} \tag{6.18}$$

for the interaction energy between delocalized charge distributions at sites A and B. Evaluating the integrals of eq 6.18 in MM–MM electrostatic energy calculations would increase the computational cost; so for low-cost calculations, one should develop efficient schemes to evaluate or approximate the integrals; for example these integrals can be replaced by point charge interactions for sites separated by more than a certain distance.

Because the optimized values of the parameters in our treatment of screened electrostatics are physical, they can be used without reoptimization in new force fields or in other methods that incorporate electrostatic effects, such as the electrostatically embedded many-body method,[70,71] the electrostatically embedded multiconfiguration molecular mechanics method,[72,73] or fragment methods for large molecular systems.[74]

The present study only considered QM–MM boundaries that pass between non bonded fragments. In future work we will consider the use of this kind of scheme when QM–MM boundaries pass through covalent bonds.

6.6 References for Chapter 6

1. *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J.; Thompson, M. A., Eds.; ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998.
2. Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*, Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; p 285.
3. Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
4. Senn, H. M.; Thiel, W. *Angew. Chemie Int. Ed.* **2009**, *48*, 1198.
5. Bernstein, N.; Kermode, J. R.; Csányi, G. *Rep. Prog. Phys.* **2009**, *72*, 026501.
6. Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
7. Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
8. Sokalski, W. A.; Poirier, R. A. *Chem. Phys. Lett.* **1983**, *98*, 86.
9. Leverentz, H.; Gao, J.; Truhlar, D. G. *Theor. Chem. Acc.*, in press.
10. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
11. Hancock, G. C.; Truhlar, D. G.; Dykstra, C. E. *J. Chem. Phys.* **1988**, *88*, 1786.
12. Stone, A. J. In *The Theory of Intermolecular Forces*; Oxford University Press Inc.: New York, 1996; p 94.

13. Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
14. Cisneros, G. A.; Tholander, S. N.-I.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J. P. *Int. J. Quantum Chem.* **2008**, *108*, 1905.
15. Bredow, T.; Geudtner, G.; Jug, K. *J. Chem. Phys.* **1996**, *105*, 6395.
16. Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
17. Yudanov, I. V.; Nasluzov, V. A.; Neyman, K. M.; Rösch, N. *Int. J. Quantum Chem.* **1997**, *65*, 975.
18. López, N.; Illas, F. *J. Phys. Chem. B* **1998**, *102*, 1430.
19. Pacchioni, G.; Ferrari, A. M. *Catal. Today* **1999**, *50*, 533.
20. Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300.
21. Soave, R.; Pacchioni, G. *Chem. Phys. Lett.* **2000**, *320*, 345.
22. Gomes, J. R. B.; Illas, F.; Hernández, N. C.; Márquez, A.; Sanz, J. F. *Phys. Rev. B* **2002**, *65*, 125414.
23. Gomes, J. R. B.; Illas, F.; Hernández, N. C.; Sanz, J. F.; Wander, A.; Harrison, N. M. *J. Chem. Phys.* **2002**, *116*, 1684.
24. Gomes, J. R. B.; Lodziana, Z.; Illas, F. *J. Phys. Chem. B* **2003**, *107*, 6411.
25. Cinquini, F.; Valentin, C. D.; Finazzi, E.; Giordano, L.; Pacchioni, G. *Theor. Chem. Acc.* **2007**, *117*, 827.
26. Valero, R.; Gomes, J. R. B.; Truhlar, D. G.; Illas, F. *J. Chem. Phys.* **2008**, *129*, 124710.
27. Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2010**, *6*, 2469.
28. Köster, A. M.; Kölle, C.; Jug, K. *J. Chem. Phys.* **1993**, *99*, 1224.
29. Kairys, V.; Jensen, J. H. *Chem. Phys. Lett.* **1999**, *315*, 140.
30. Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353.
31. Hall, G. G.; Smith, C. M. *Int. J. Quantum Chem.* **1992**, *42*, 1237.
32. Wheatley, R. J.; Mitchell, J. B. O. *J. Comput. Chem.* **1994**, *15*, 1187.
33. Guillot, B.; Guissani, Y. *J. Chem. Phys.* **2001**, *114*, 6720.

34. Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
35. Piquemal, J. P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.
36. Torheyden, M.; Jansen, G. *Mol. Phys.* **2006**, *104*, 2101.
37. Spackman, M. A. *Chem. Phys. Lett.* **2006**, *418*, 158.
38. Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Phys. Chem. B* **2006**, *110*, 13682.
39. Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
40. Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190.
41. Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309.
42. Slater, J. C. *Phys. Rev.* **1930**, *36*, 57.
43. Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
44. Clementi, E.; Raimondi, D. L. *J. Chem. Phys.* **1963**, *38*, 2686.
45. Cusachs, L. C.; Trus, B. L.; Carroll, D. G.; Mcglynn, S. P. *Int. J. Quantum Chem. Symp.* **1967**, *1*, 423.
46. Massey, H. S. W.; Burhop, E. H. S. In *The international series of monographs on physics-Electronic and ionic impact phenomena*, Marshall, W., Wilkinson, D. H., Eds.; University Press: Oxford, 1969; p 376.
47. Archambault, F.; Chipot, C.; Soteras, I.; Luque, F. J.; Schulten, K.; Dehez, F. *J. Chem. Theory Comput.* **2009**, *5*, 3022.
48. Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
49. Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
50. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
51. Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
52. Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S. A.; Lotrich, V. F.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Podeszwa, R.; Rybak, S.; Szalewicz, K.; Williams, H. L.; Wheatley, R. J.; Wormer, P. E. S.; Żuchowski, P. S. *SAPT2008: "An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies" version* 2008.

53. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, version D. 01*; Gaussian, Inc.: Wallingford, CT, 2004.
54. Ponder, J. W. *TINKER, version 4.2*; Washington University: St. Louis, MO, 2004.
55. Lin, H.; Zhang, Y.; Truhlar, D. G. *QMMM, version 1.3.5*; University of Minnesota: Minneapolis, 2007.
56. Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.
57. Dahlke, E. E.; Orthmeyer, M. A.; Truhlar, D. G. *J. Phys. Chem. B* **2008**, *112*, 2372.
58. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
59. Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
60. Curtiss, L. A.; McGrath, M. P.; Blaudeau, J. P.; Davis, N. E.; Binning, R. C.; Radom, L. *J. Chem. Phys.* **1995**, *103*, 6104.
61. Stewart, R. F. *J. Chem. Phys.* **1969**, *50*, 2485.
62. Szabo, A.; Ostlund, N. S. In *Modern Quantum Chemistry-Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: Mineola, New York, 1989; p 157.
63. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.;

Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09, version A. 02*; Gaussian, Inc.: Wallingford CT, 2009.

64. Strand, T. G.; Bonham, R. A. *J. Chem. Phys.* **1964**, *40*, 1686.
65. Reinhardt, P.; Piquemal, J. P. *Int. J. Quantum Chem.* **2009**, *109*, 3259.
66. Gill, P. M. W. *J. Phys. Chem.* **1996**, *100*, 15421.
67. Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1988**, *89*, 5739.
68. Nygren, M. A.; Pettersson, L. G. M.; Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1994**, *100*, 2010.
69. Rotenberg, B.; Salanne, M.; Simon, C.; Vuilleumier, R. *Phys. Rev. Lett.* **2010**, *104*, 138301.
70. Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
71. Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1.
72. Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 790.
73. Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2925.
74. Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. *J. Phys. Chem. B* **2009**, *113*, 9646.

Table 6.1 Clementi–Raimondi exponents for the outermost orbitals[a]

Atom	H	C	N	O	F
exponent	1.00	1.57	1.92	2.23	2.55
Atom	Si	P	S	Cl	Br
exponent	1.43	1.63	1.83	2.04	2.26

[a] Ref. 44

Table 6.2 The contraction coefficients and exponents[a]

	C_i	α_i
1	0.107150	0.109818
2	0.343808	0.405771
3	0.355483	2.227660

[a] Refs. 61 and 62

Table 6.3 Averaged SAPT electrostatic, induction, and damped induction energies (kcal/mol) using three geometries and two basis sets.

	equilibrium /acTZ[a]	equilibrium /def2-TZVP	compressed /acTZ	compressed /def2-TZVP	extended /acTZ	extended /def2-TZVP	All
Electrostatic	–14.50	–18.78	–25.58	–30.89	–9.20	–12.80	–18.98
Induction	–9.61	–9.58	–24.86	–25.14	–4.14	–4.08	–12.91
Damped-induction	–3.18	–3.49	–7.87	–8.58	–1.63	–1.77	–4.45

[a] acTZ represents aug-cc-pVTZ

Table 6.4 Averaged SAPT electrostatic, induction, and damped induction energies (kcal/mol) using three geometries and the def2-TZVP basis set over only 29 molecules.

	equilibrium /def2-TZVP	compressed /def2-TZVP	extended /def2-TZVP
Electrostatic	–15.17	–26.32	–9.80
Induction	–9.55	–24.82	–4.08
Damped-induction	–3.07	–7.73	–1.53

Table 6.5 Zeta values used in the Slater-type orbital

Atom	H	C	N	O	F
optimized parameters	1.32	0.92	0.92	1.20	1.16
Strand and Bonham[a]	1.00	0.87	1.01	1.12	1.24
Atom	Si	P	S	Cl	Br
optimized parameters	0.73	0.68	0.90	0.98	0.91
Strand and Bonham[a]	0.74	0.81	0.88	0.95	1.01

[a] half of the values in Ref. 64

Table 6.6 MSE and MUE of electrostatic energies (kcal/mol) using the QM/MM method. (Exact values are SAPT electrostatic energies)

	equilibrium /acTZ		equilibrium /def2-TZVP		compressed /acTZ		compressed /def2-TZVP		extended /acTZ		extended /def2-TZVP		All	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
Pt charge[a]	6.43	6.43	6.43	6.43	14.80	14.80	15.10	15.10	3.02	3.02	2.98	2.98	8.13	8.13
Opt[b]	1.58	1.97	1.32	2.17	3.64	5.09	2.61	5.31	1.06	1.11	1.03	1.16	1.84	2.81
MSB[c]	1.59	2.22	1.19	2.29	3.55	5.39	2.15	5.90	1.07	1.25	1.00	1.24	1.71	3.06

[a] point charge scheme

[b] screened charge scheme with optimized parameters

[c] screened charge scheme with modified Strand-Bonham (MSB) parameters

Table 6.7 MSE and MUE of the induction energies (kcal/mol) using the QM/MM method. (Exact values are SAPT damped induction energies)

	equilibrium /acTZ		equilibrium /def2-TZVP		compressed /acTZ		compressed /def2-TZVP		extended /acTZ		extended /def2-TZVP		All	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
Pt charge[a]	0.67	1.09	1.12	1.18	3.46	3.98	4.34	4.46	0.12	0.45	0.35	0.38	1.72	1.94
Opt[b]	–0.74	1.00	0.69	0.82	–0.62	2.67	2.68	3.13	–0.34	0.45	0.25	0.30	0.46	1.40
MSB[c]	–0.79	1.02	0.68	0.80	–0.71	2.65	2.60	3.12	–0.36	0.46	0.25	0.29	0.42	1.39

[a] point charge scheme

[b] screened charge scheme with optimized parameters

[c] screened charge scheme with modified Strand-Bonham (MSB) parameters

Table 6.8 Electrostatic and induction energies (kcal/mol) of $HCONH_2$…H_2O dimer in QM/MM calculations compared with SAPT results, and MUE (kcal/mol) of QM/MM calculations over three geometries and two basis sets.

		equilibrium /acTZ	equilibrium /def2-TZVP	compressed /acTZ	compressed /def2-TZVP	extended /acTZ	extended /def2-TZVP	MUE
Electrostatic								
SAPT		–11.3	–11.6	–21.9	–22.3	–6.6	–6.9	
QM/MM	Pt charge	–7.4 / –7.3[a]	–8.0 / –7.5	–10.5 / –11.3	–11.4 / –11.5	–5.1 / –5.0	–5.5 / –5.1	5.4
QM/MM	Opt	–10.5 / –9.8	–10.8 / –10.0	–22.0 / –18.6	–22.5 / –18.8	–5.9 / –5.7	–6.2 / –5.8	1.3
QM/MM	MSB	–10.8 / –10.5	–11.2 / –10.7	–23.4 / –19.5	–23.8 / –19.8	–6.0 / –6.0	–6.3 / –6.1	1.1
Induction								
SAPT (damped)		–1.7	–1.6	–4.6	–4.5	–0.7	–0.7	
QM/MM	Pt charge	–1.2	–1.1	–2.8	–2.5	–0.6	–0.5	0.9
QM/MM	Opt	–2.0	–1.3	–6.3	–3.7	–0.7	–0.5	0.5
QM/MM	MSB	–2.1	–1.3	–6.8	–3.9	–0.8	–0.5	0.6

[a] *x*/*y* denotes that the electrostatic energy is *x* when $HCONH_2$ is the QM region and is *y* when H_2O is the QM region.

Table 6.9 Electrostatic and induction energies (kcal/mol) of HSO_4^-…NH_4^+…H_2O in QM/MM calculations compared with SAPT results, and MUE (kcal/mol) of QM/MM calculations over three geometries.

		equilibrium /def2-TZVP	compressed /def2-TZVP	extended /def2-TZVP	MUE
Electrostatic					
SAPT		–144.1	–176.4	–124.0	
QM/MM	Pt charge	–126.4 / –130.2[a]	–135.0 / –147.6	–116.4 / –117.5	19.3
QM/MM	Opt	–149.4 / –138.7	–190.9 / –161.5	–124.7 / –121.6	7.2
QM/MM	MSB	–146.0 / –139.4	–184.6 / –162.2	–123.2 / –122.3	5.2
Induction					
SAPT (damped)		–15.7	–30.6	–9.8	
QM/MM	Pt charge	–12.2	–17.4	–8.8	5.9
QM/MM	Opt	–15.7	–26.7	–9.9	1.4
QM/MM	MSB	–14.9	–25.4	–9.5	2.1

[a] *x*/*y* denotes that the electrostatic energy is *x* when HSO_4^- is the QM region and is *y* when NH_4^+ (H_2O) is the QM region.

Table 6.10 Electrostatic and induction energies (kcal/mol) of HCl…H_2S dimer in QM/MM calculations compared with SAPT results, and MUE (kcal/mol) of QM/MM calculations over three geometries and two basis sets.

		equilibrium /acTZ	equilibrium /def2-TZVP	compressed /acTZ	compressed /def2-TZVP	extended /acTZ	extended /def2-TZVP	MUE
Electrostatic								
SAPT		–5.2	–5.4	–10.9	–11.1	–2.8	–3.0	
QM/MM	Pt charge	–1.0 / –2.1[a]	–1.1 / –2.4	–1.7 / –2.7	–1.8 / –3.3	–0.6 / –1.5	–0.7 / –1.7	4.7
QM/MM	Opt	–2.4 / –4.5	–2.5 / –4.7	–4.0 / –10.5	–4.0 / –11.1	–1.2 / –2.2	–1.2 / –2.3	2.2
QM/MM	MSB	–2.5 / –4.6	–2.5 / –4.9	–3.9 / –11.0	–3.9 / –11.6	–1.2 / –2.2	–1.3 / –2.4	2.2
Induction								
SAPT (damped)		–1.1	–1.0	–3.0	–3.0	–0.5	–0.4	
QM/MM	Pt charge	–0.4	–0.3	–0.8	–0.6	–0.2	–0.1	1.1
QM/MM	Opt	–1.5	–0.5	–5.1	–1.9	–0.4	–0.2	0.7
QM/MM	MSB	–1.5	–0.5	–5.4	–2.0	–0.5	–0.2	0.8

[a] *x*/*y* denotes that the electrostatic energy is *x* when HCl is the QM region and is *y* when H_2S is the QM region.

Table 6.11 MUE (kcal/mol) of QM/MM electrostatic energies and improvement ratios (IR) for individual elements.

Element	H	C	N	O	F	Si	P	S	Cl	Br
First division scheme										
number of calculations[a]	372	78	87	171	54	6	12	54	60	6
Pt charge	7.9	6.9	10.2	7.5	9.6	11.0	4.9	7.3	13.7	3.6
Opt	2.9	2.4	5.1	2.6	1.6	1.7	2.9	2.5	2.1	1.4
MSB	3.1	2.7	5.5	2.7	2.0	2.2	3.1	2.6	2.6	1.6
IR (Opt)[b]	2.6	2.6	1.9	2.8	4.7	4.9	1.6	2.8	5.3	2.3
Second division scheme										
number of calculations[a]	213	12	66	96	18	0	12	18	30	6
Pt charge	6.9	5.2	11.1	6.1	5.9		4.9	4.0	18.6	3.6
Opt	1.8	0.7	5.7	2.8	1.2		2.9	2.3	2.7	1.4
MSB	1.8	1.2	6.6	3.0	2.1		3.1	2.2	3.1	1.6
IR (Opt)	3.8	7.7	1.9	2.1	4.9		1.7	1.7	6.9	2.6

[a] This number includes all calculations using different geometries and different basis sets.

[b] Improvement ratios (IRs) for individual elements using the screened charge scheme with optimized parameters

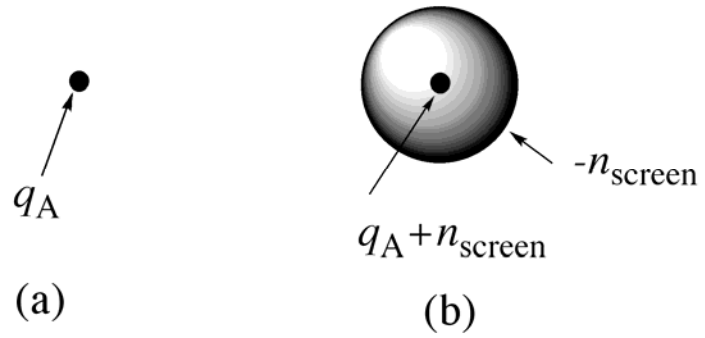


Figure 6.1 Comparison between (a) a point charge model and (b) a screened charge model of an MM atom A. The total smeared charge in model (b) is $-n_{screen}$, representing n_{screen} electrons.

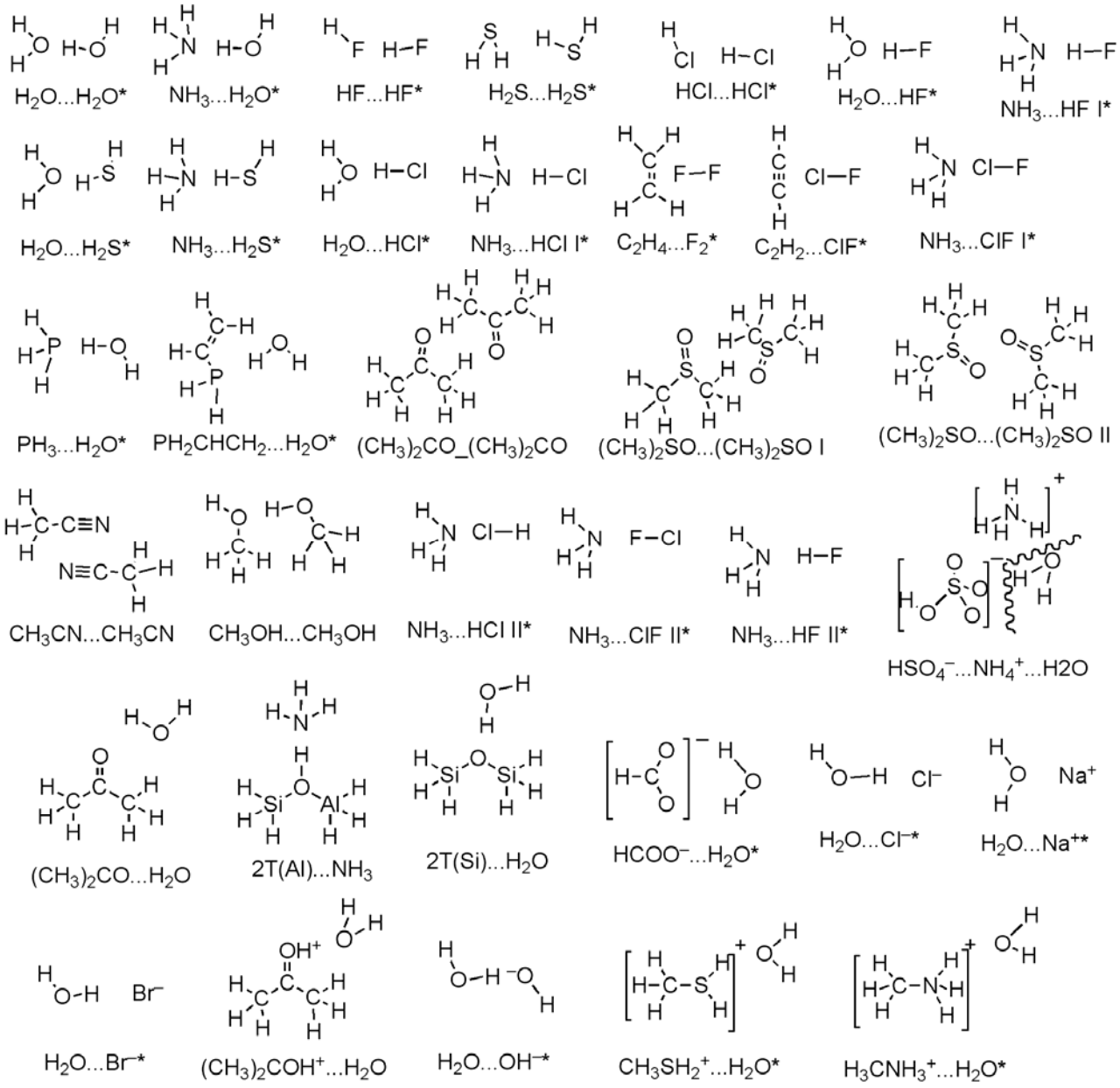


Figure 6.2 36 of 40 dimers in the test suite. We use * to label the molecules that are tested using both aug-cc-pVTZ and def2-TZVP basis sets.

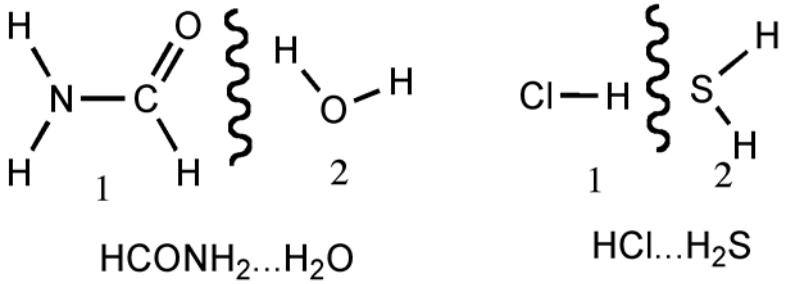


Figure 6.3 Geometry of the 37th and 38th dimers.

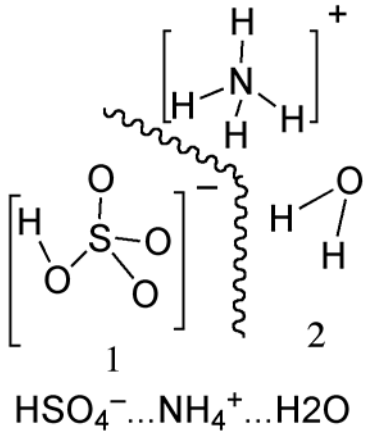


Figure 6.4 Geometry of the 39th dimer.

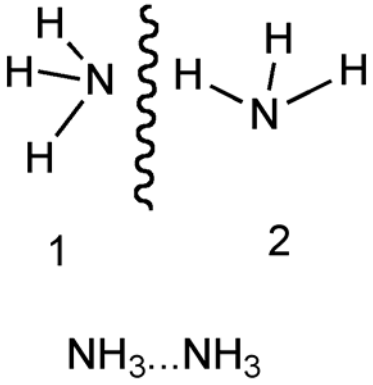


Figure 6.5 Geometry of the 40th dimer.

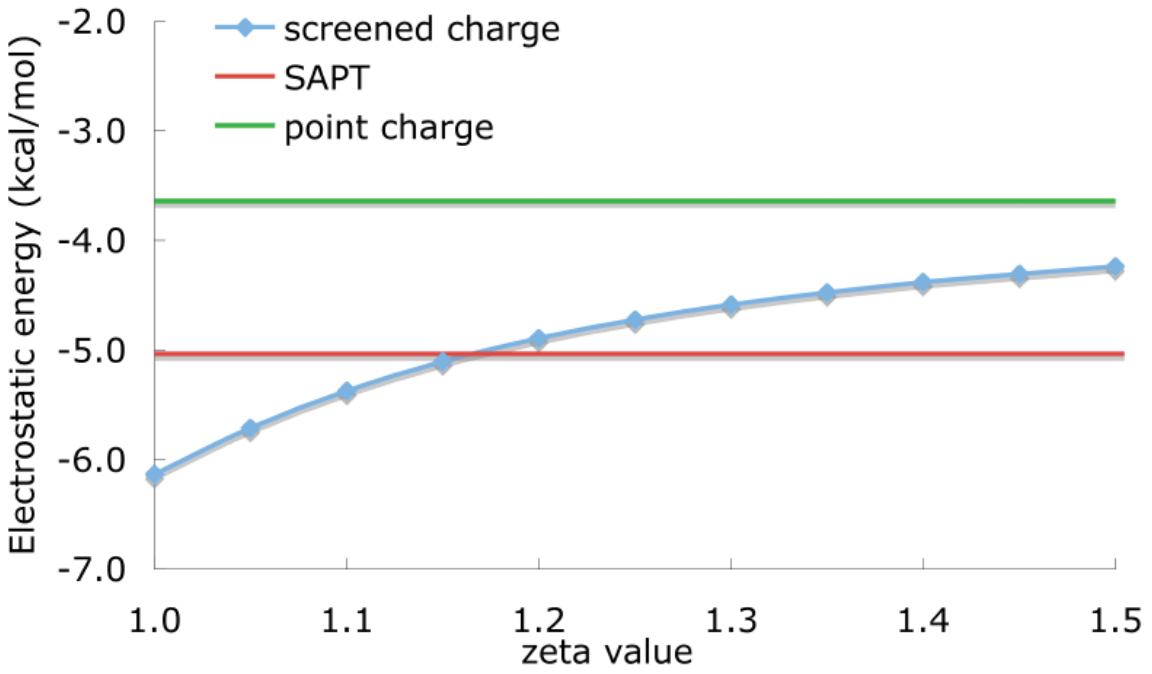


Figure 6.6 QM/MM and SAPT electrostatic energies (kcal/mol) with respect to the H zeta value.

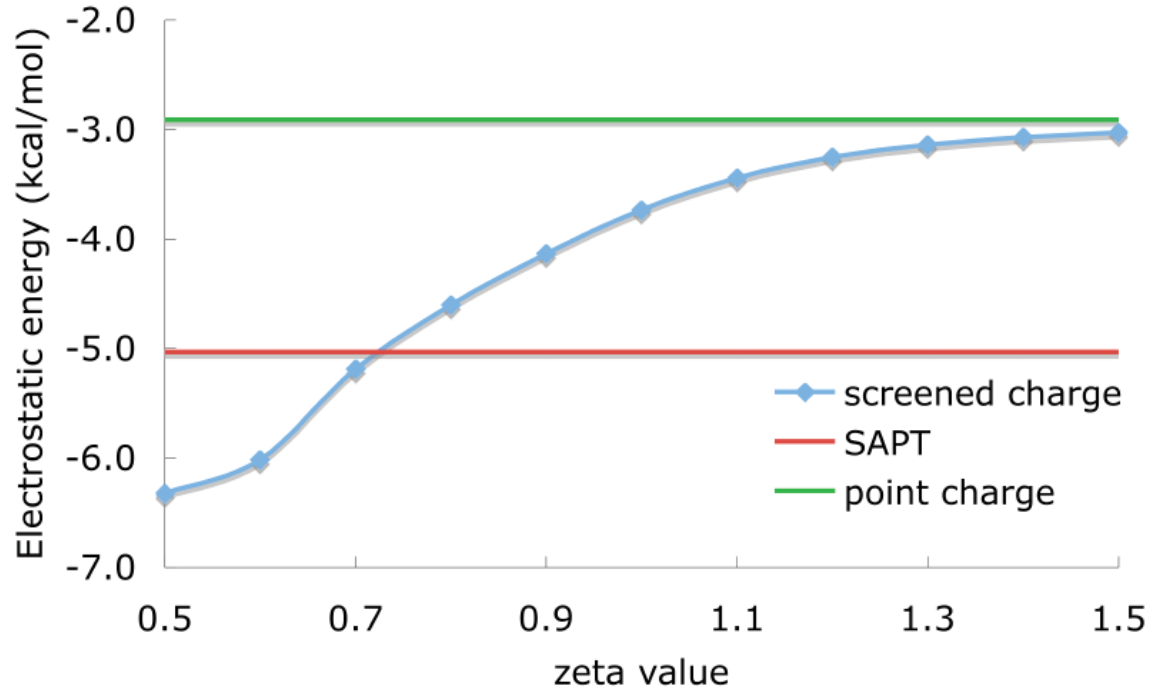


Figure 6.7 QM/MM and SAPT electrostatic energies (kcal/mol) with respect to the N zeta value.

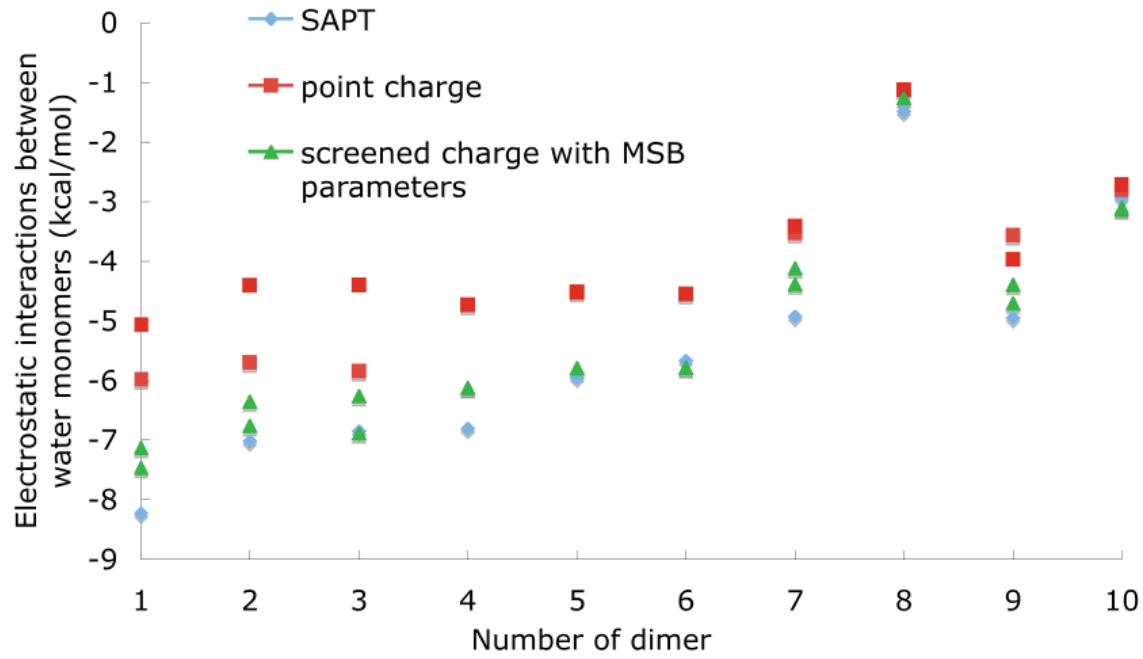


Figure 6.8 QM/MM and SAPT electrostatic energies (kcal/mol) for ten water dimers.

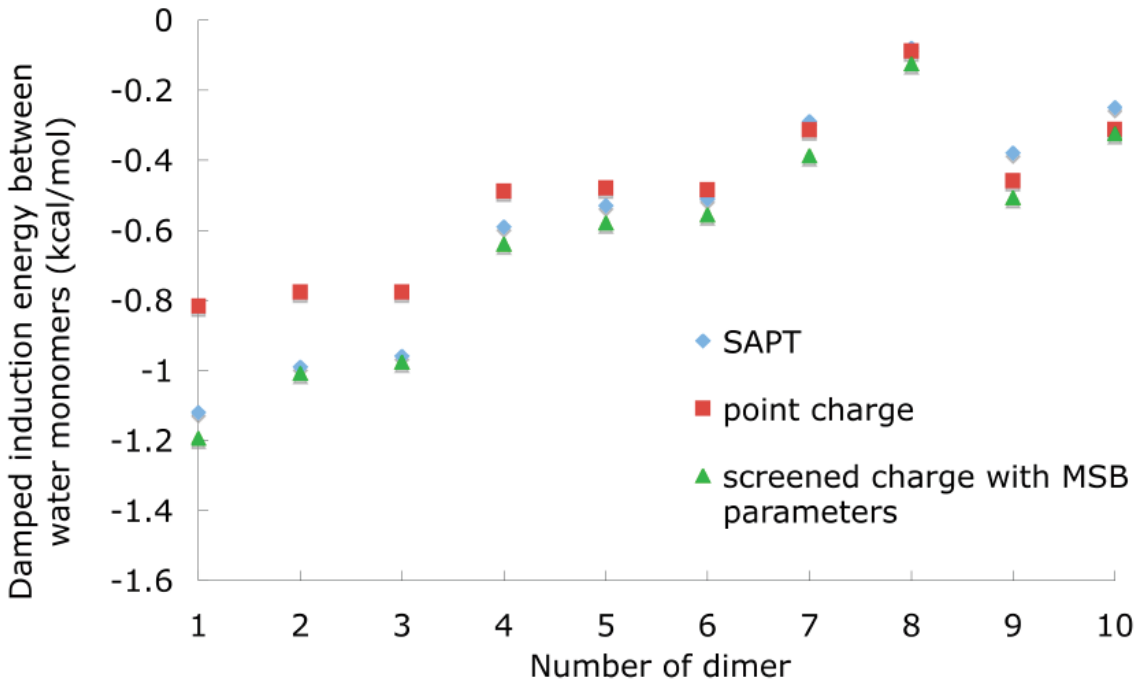


Figure 6.9 QM/MM induction energies and SAPT damped induction energies (kcal/mol) for ten water dimers.

Chapter 7. Screened Electrostatically Embedded Many-Body Method[6]

7.1 Introduction

One aim of modern quantum chemistry research is the development of methods that enable more accurate calculations at lower computational costs for large or complex systems. Numerous types of fragmentation-based approaches have been investigated as approximations that reduce overall computational cost while maintaining a reasonable degree of accuracy[1-11]. One particularly promising fragmentation-based method is the electrostatically embedded many-body (EE-MB) method,[5,12-17] and the present letter shows how this method can be improved.

The EE-MB method has been presented elsewhere,[4] so only the necessary essentials are reviewed here. In the many-body approach, the system of interest is first partitioned into *N* fragments called monomers. The system energy is written as a finite series of *N* terms as

$$E_{\text{sys}} = V_1 + V_2 + ... + V_N \tag{7.1}$$

where

[6] This work was supported in part by the National Science Foundation under Grant Nos. CHE09-56776 and CHE10-51396 and by the Lando/NSF summer undergraduate research program of the Department of Chemistry of the University of Minnesota.

$$V_1 = \sum_{i=1}^{N} E_i \tag{7.2}$$

$$V_2 = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (E_{ij} - E_i - E_j) \tag{7.3}$$

$$V_3 = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^{N} \left[E_{ijk} - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) - (E_{jk} - E_j - E_k) - E_i - E_j - E_k \right] \tag{7.4}$$

and so forth; $E_{i\ldots k}$ is the single-point energy of the fragment containing monomers *i* through *k*. Approximating the system energy by truncating eq 7.1 at the V_2 term gives the pairwise additive (PA) approximation while truncating eq 7.1 at the V_3 term gives the three-body (3B) approximation.

In previous work, embedding the monomer, dimer, and trimer fragments in an electrostatic potential[18] defined by a series of point charges centered at the nuclei of the atoms external to the fragment, each with a magnitude equal to the partial atomic charge q_a of the corresponding atom *a*, was shown to increase the accuracy of the PA and 3B calculations.[4,12,13] These electrostatically embedded many-body calculations are labeled EE-PA and EE-3B.

In other contexts, better approximations to electrostatic potentials have been obtained by using methods such as distributed multipoles,[19-21] off-nuclei point charges,[22,23] and penetration effects.[24-28] Here we explore the effect of using a better approximation to the electrostatic potential in the EE-MB method; in particular, we

examine the use of a recently developed[28] screened charge model to include penetration effects. In this model, the outermost portion of the charge density of an atom is distributed in a Slater type orbital (STO), and the new effective charge of atom *a* at a point in space a distance r_a from it can be expressed as

$$q_a^* = q_a + n_{\text{screen},a} f_{Z_a}(\zeta_{Z_a} r_a) \exp(-2\zeta_{Z_a} r_a) \tag{7.5}$$

where f_{Z_a} is a polynomial[28] depending on the atomic number Z_a of atom *a*, ζ_{Z_a} is the zeta parameter of the STO, and $n_{\text{screen},a}$ is the number of electrons in the STO. In this letter, we examine the effect of implementing the screened charge model in the EE-MB method at both the PA and 3B levels, and the resulting new models are called screened EE-MB methods, in particular sEE-PA and sEE-3B. The difference in cost for unscreened and screened calculations is negligible since the difference affects only the relatively inexpensive one-electron-integral portion of the calculation.

7.2 Methods

The methods are tested on the binding energies of a water trimer and five water hexamers, whose structures are shown in Fig. 1. One of the advantages of the PA and 3B methods is that their computational effort scales as N^2 and N^3, respectively, for large *N*. Thus, for example, if a standard calculation scales as N^7, these methods lower the scaling for large systems dramatically. For testing the methods, though, we here use density functional theory (DFT) rather than an N^7 method. The test consists of performing an unfragmented DFT calculation on the entire system (for a hexamer, this is a 6B

calculation) and using that as reference energy. For the hexamers, we compute the mean unsigned deviation (MUD) of the various PA and 3B calculations from the reference 6B value (where the deviation is averaged over the five structures), and we use this MUD as a measure of the accuracy of each fragmentation scheme. The water trimer, called C1GM, is the MP2/aug-cc-pVQZ global minimum structure taken from previous work.[29] The five water hexamers, labeled boat, book, cage, prism, and ring, are the optimized geometries taken from Day et al.,[30] as used previously.[13]

For each type of calculation, the binding energy of the water structure was calculated as

$$E_{\text{bind}} = \sum_{i=1}^{N} E_i' - E_{\text{sys}} \tag{7.6}$$

where E_i' is the single-point energy of monomer *i* at the geometry that the monomer has in the cluster, *N* is the total number of fragments, and E_{sys} is the single-point energy of the system. In each structure, a single water molecule constitutes a single fragment. Hence for the C1GM structure, *N* equals 3, while for the five hexamers, *N* is 6. For a trimer, the PA approximation requires 3 dimer calculations. For a hexamer, the PA approximation requires 15 dimer calculations, and the 3B calculation requires, in addition, 20 trimer calculations.

All DFT calculations were performed using the M06-2X density functional[31] with the 6-311+G(2df,2p) basis set. Conventional calculations were performed using *Gaussian 09,*[32] and EE-MB and sEE-MB calculations were performed using the

MBPAC 2011-3 software package,[33] which is a front end for *Gaussian 09* and enables the automation of the calculations. Values for partial atomic charges (which are taken to be independent of geometry in the embedding calculations) were determined by the Merz-Kollman (MK) algorithm[34] for a DFT-optimized H_2O structure. This yielded values of -0.727364 *e* for the O atoms and 0.363682 *e* for the H atoms, where *e* is the charge on a proton.

The parameters for electrostatic screening are all taken from the original electrostatic screening paper.[28] In particular, ζ_{Z_a} is 1.20 and 1.32 bohr^{-1} for H and O, respectively, and $n_{\text{screen},a}$ is 1.0 for O atoms and $1-q_a$ for H atoms.

7.3 Results and discussion

The unfragmented DFT calculation resulted in an E_{bind} value of 18.13 kcal/mol for the C1GM trimer. The MB calculations resulted in E_{bind} values of 15.69 kcal/mol for the PA calculation, 17.73 kcal/mol for the EE-PA calculation, and 17.97 kcal/mol for the screened EE-PA calculation. The improvement in these results as the electrostatics are approximated more accurately is very encouraging. Note that all three 3B schemes would be exact for a trimer, so we consider the hexamers for a more challenging test.

The results for the hexamers are in Table 7.1, which shows not only E_{bind} and MUD but also the mean unsigned percent deviation (MUPD) relative to the reference 6B result. The table shows that the utilization of screened charges in the MB calculations results in a dramatic improvement in the accuracy of the E_{bind} values relative to the

accuracies attained with unscreened charges. The sEE-PA method shows an MUD of only 0.60 kcal/mol, which is only about 1% of the binding energy. This compares favorably to the PA value of 1.32 kcal/mol (2.6%) obtained with unscreened charges and is even on par with the EE-3B MUD of 0.54 kcal/mol (1.0%). This indicates that the screened charge pairwise model has the ability to achieve a level of accuracy comparable to that obtained at the three-body level with unscreened charges! Considering the much smaller cost and better scaling of the pairwise additive calculation, this is a very significant finding.

Furthermore, improvements in accuracy of the MB method are also seen at the 3B level when screened charges are implemented. In particular, the sEE-3B MUD of 0.24 kcal/mol (0.5%) is more than a factor of two smaller than the EE-3B MUD of 0.54 kcal/mol (1.0%) and is a factor of 7 smaller than the 3B MUD of 1.78 kcal/mol (3.5%).

7.4 Conclusions

The results for the $(H_2O)_6$ structures examined here indicate that a very high level of accuracy can be obtained with an sEE-PA calculation. This suggests the possibility for highly accurate calculations for much larger clusters. The relative speed that may be achieved with PA calculations (as compared to the 3B calculations) also suggests that Monte Carlo dynamics simulations of large clusters may be possible using the sEE-MB model. One advantage of the present screening algorithm is that the electrostatic potential of the charged sites is a radial function that allows analytic gradients with a negligible increase in complexity as compared to the analytic gradients

that have already been implemented[17] for the unscreened EE-MB method; thus the method can also be useful for molecular dynamics.

The improvements seen here with screened electrostatics also indicate that other approximation methods employing electrostatic potentials would also benefit from this improved electrostatics methodology.

7.5 References for Chapter 7

(1) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys*. **2003**, *119*, 3599-3606.

(2) Fedorov, D. G.; Kitaura, K. The Importance of Three-Body Terms in the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *120*, 6832-6840.

(3) Li, S.; Li, W.; Fang, T. An Efficient Fragment-Based Approach for Predicting the Ground-State Energies and Structures of Large Molecules. *J. Am. Chem. Soc*. **2005**, *127*, 7215-7226.

(4) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters. *J. Chem. Theory Comput*. **2007**, *3*, 46-53.

(5) Collins. M. A.; Deev, V. A. Accuracy and Efficiency of Electronic Energies from Systematic Molecular Fragmentation. *J. Chem. Phys*. **2006**, *125*, 104104-104119.

(6) Fedorov, D. G.; Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A.* **2007**, *111*, 6904-6914.

(7) Hirata, S.; Yagi, K. Predictive Electronic and Vibrational Many-Body Methods for Molecules and Macromolecules. *Chem. Phys. Lett*. **2008**, *464*, 123-134.

(8) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. The Variational Explicit Polarization Potential and Analytical First Derivative of Energy: Towards a Next Generation Force Field. *J. Chem. Phys*. **2008**, *128*, 234108-234117.
(9) Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B*. **2009**, *113*, 9646-9663.
(10) Söderhjelm, P.; Aquilante, F.; Ryde, U. Calculation of Protein-Ligand Interaction Energies by a Fragmentation Approach Combining High-Level Quantum Chemistry with Classical Many-Body Effects. *J. Phys. Chem*. *B*. **2009**, *113*, 11085-11094.
(11) Li, W.; Piecuch, P. Multilevel Extension of the Cluster-in-Molecule Local Correlation Methodology: Merging Coupled-Cluster and Møller-Plesset Perturbation Theories. *J. Phys. Chem. A*. **2010**, *114*, 6721-6727.
(12) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Correlation Energy, with Applications to the Calculation of Accurate Second-Order Møller-Plesset Perturbation Theory Energies for Large Water Clusters. *J. Chem. Theory Comput*. **2007**, *3*, 1342-1348.
(13) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. Evaluation of the Electrostatically Embedded Many-Body Expansion and the Electrostatically Embedded Many-Body Expansion of the Correlation Energy by Application to Low-Lying Water Hexamers. *J. Chem. Theory Comput*. **2008**, *4*, 33-41.
(14) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Simulations. *J. Chem. Theory Comput*. **2008**, *4*, 1-6.
(15) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. Application of the Electrostatically Embedded Many-Body Expansion to Microsolvation of Ammonia in Water Clusters. *J. Chem. Theory Comput*. **2008**, *4*, 683-688.
(16) Leverentz, H. R.; Truhlar, D. G. Electrostatically Embedded Many-Body Approximation for Systems of Water, Ammonia, and Sulfuric Acid and the Dependence of Its Performance on Embedding Charges. *J. Chem. Theory Comput*. **2009**, *5*, 1573-1584.

(17) Speetzen, E. D.; Leverentz, H. R.; Lin, H.; Truhlar, D. G. *Accurate Condensed Phase Electronic Structure Theory*; Manby, F., Ed.; CRC Press: Boca Raton, 2010.
(18) Politzer, P., Truhlar, D. G., Eds.; *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Plenum: New York, 1981.
(19) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233-239.
(20) Sokalski, W. A.; Poirier, R. A. Cumulative Atomic Multipole Representation of the Molecular Charge Distribution and Its Basis Set Dependence. *Chem. Phys. Lett.* **1983**, *98*, 86.
(21) Leverentz, H.; Gao, J.; Truhlar, D. G. Using Multipole Point Charge Distributions to Provide the Electrostatic Potential in the Variational Explicit Polarization (X-Pol) Potential. *Theor. Chem. Acc.* **2011**, *129*, 3-13.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
(23) Hancock, G. C.; Truhlar, D. G.; Dykstra, C. E. An Analytic Representation of the Six-dimensional Potential Energy Surface of Hydrogen Fluoride Dimer. *J. Chem. Phys.* **1988**, *88*, 1786-1796.
(24) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press Inc.: New York, 1996.
(25) Kairys, V.; Jensen, J. H. Evaluation of the Charge Penetration Energy Between Non-orthogonal Molecular Orbitals Using the Spherical Gaussian Overlap Approximation. *Chem. Phys. Lett.* **1999**, *315*, 140-144.
(26) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. Evaluation of Charge Penetration Between Distributed Multipolar Expansions. *J. Chem. Phys.* **2000**, *112*, 7300-7306.
(27) Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C. Improved Formulas for the Calculation of the Electrostatic Contribution to the Intermolecular Interaction Energy

from Multipolar Expansion of the Electronic Distribution. *J. Phys. Chem. A.* **2003**, *107*, 10353-10359.

(28) Wang, B.; Truhlar, D. G. Including Charge Penetration Effects in Molecular Modeling. *J. Chem. Theory Comput.* **2010**, *6*, 3330-3342.

(29) Dahlke, E. E.; Truhlar, D. G. Assessment of Pairwise Additive Approximation and Evaluation of Many-Body Terms for Water Clusters. *J. Phys. Chem. B* **2006**, *110*, 10595-10601.

(30) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. A study of water clusters using the effective fragment potential and Monte Carlo simulated annealing. *J. Chem. Phys.* **2000**, *112*, 2063-2073.

(31) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09 Revision A.01*; Gaussian, Inc.: Wallingford, CT, 2009.

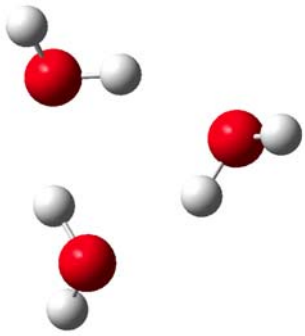
(33) Dahlke, E. E.; Lin, H.; Leverentz, H.; Wang, B.; Truhlar, D. G. *MBPAC 2011-3*; University of Minnesota: Minneapolis, MN, 2011.

(34) Besler, B. H.; Merz Jr., K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comp. Chem.* **1990**, *11*, 431-439.

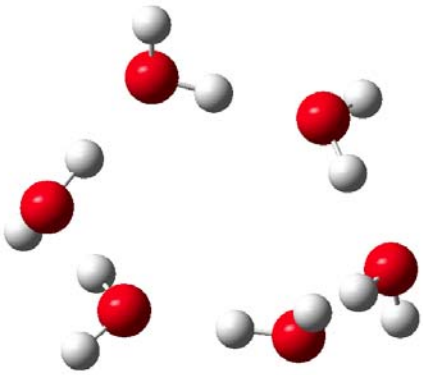
Table 7.1 Binding energy (E_{bind} in kcal/mol), mean unsigned deviation (MUD in kcal/mol) in E_{bind}, and mean unsigned percent deviation (MUPD) in E_{bind}

Structure	6B	PA	3B	EE-PA	EE-3B	sEE-PA	sEE-3B
Boat	48.52	35.55	46.44	47.10	48.11	48.42	48.33
Book	51.11	40.41	49.52	50.06	50.58	51.81	50.84
Cage	53.34	43.94	52.20	52.00	52.85	54.20	53.25
Prism	54.50	45.42	52.63	53.33	53.69	55.66	54.07
Ring	49.82	36.42	47.59	48.21	49.36	49.62	49.59
MUD	0[a]	11.11	1.78	1.32	0.54	0.60	0.24
MUPD	0[a]	21.8	3.5	2.6	1.0	1.1	0.5

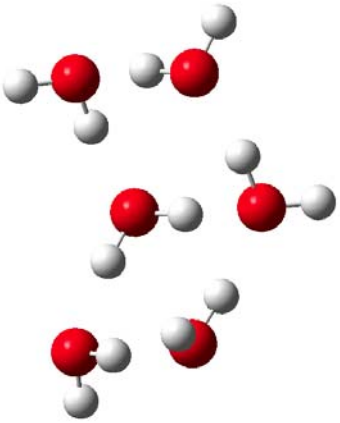
[a] by definition



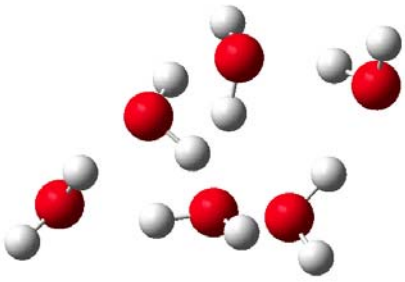
C1GM



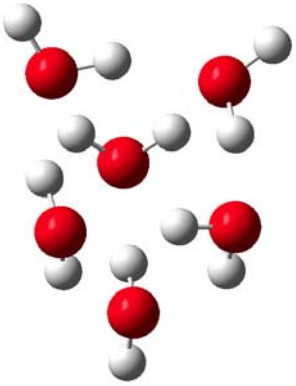
Boat



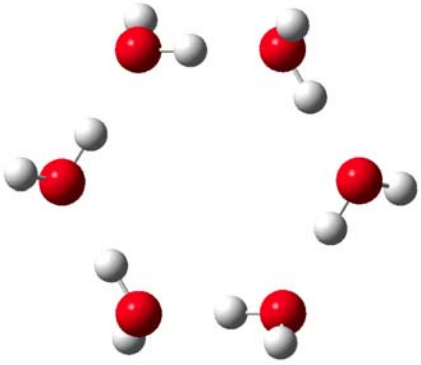
Book



Cage



Prism



Ring

Figure 7.1 Structures of the five hexamers and C1GM trimer

Chapter 8. Partial Atomic Charges and Screened Charge Models of the Electrostatic Potential[7]

8.1 Introduction

Partial atomic charges play an important role in molecular simulations. They constitute a very convenient way to describe the charge distribution within a molecule, and they can be used to compute the noncovalent electrostatic interactions within or among molecules in molecular mechanics. Moreover, the partial atomic charges are useful as an analysis tool that reflects changes in an atom's local chemical environment and bonding.

Partial atomic charge is not a quantum mechanical observable, and it is ambiguous to assign a certain portion of the electron density to each atom in a molecule; as a consequence, various charge models have been proposed to determine the partial atomic charges. They can be classified into four categories.[1] Class I charges are based on concepts and formulas originating in classical mechanical models, such as classical electronegativity equalization.[2] Class II charges are based on partitioning of the wave function or the electron density from quantum mechanical calculations; examples include Mulliken population analysis,[3] Löwdin population analysis,[4] and Hirshfeld population analysis.[5-7] Such partitions are intrinsically ambiguous. Class III charges are computed by

[7] This work was supported by the National Science Foundation under grant no. CHE09-56776.

constraining the charges to reproduce, usually in a least-squares sense, calculated physical observables, such as the electrostatic potentials.[8-16] Class IV charges are derived by semiempirical mapping of the partial atomic charges from other charge models (generally of class II) to reproduce experimental observables.[1,17-19]

Electrostatic-potential-fitted (ESP fitted) charges, which are fitted to reproduce quantum mechanically calculated electrostatic potentials, belong to Class III. The method was first developed by Momany,[8] followed by Cox and Williams,[9] Singh and Kollman,[10] Chirlian and Francl,[11] Breneman and Wiberg,[12] and others. The most significant difference among various schemes is the selection of fitting points. The ESP charges can reproduce the electrostatic potentials far from the molecule quite well and near the molecule reasonably well, and they are widely used to treat the noncovalent intramolecular and intermolecular electrostatic interactions in molecular modeling. They also lead to reasonably accurate values for some physical observables, such as the dipole moment. Moreover, compared with Class II charges, ESP charges for many molecules show smaller variations with the quantum mechanical level of theory and the basis set, and they converge to definite values with increasing basis-set size, whereas population analysis charges can and often do become unphysical for large basis sets.

Conventional ESP charges face some challenges. First, the point-charge approximation breaks down when the fitted points are close to the centers of atoms, for example, when they are closer than about one van der Waals radius. Therefore we cannot fit the points that are too close to the centers of the atoms in the fitting process, and the resulting fitting points are far from buried atoms, that is, from the atoms that are not near

the molecular surface. As a consequence, the charges of buried atoms are poorly determined. Mathematically this shows up as ill-conditioning in the linear equations of the least-squares ESP fitting process.[20] Second, the fitted charges are not numerically stable with respect to the molecular geometries. For example, unphysical conformational dependence has been found in electrostatic fitting.[10,12] This causes problems in the calculations of intramolecular interactions. One of the most widely used approaches to circumvent these problems is the restrained ESP (RESP) method, in which a penalty function is added for certain atoms in the fitting procedure to restrain their charges from deviating widely from target values. The target charges can be zero,[13] Hirshfeld charges,[14] or Bader charges.[16] However, RESP charges can also suffer some stability problems.[15]

One of the intrinsic problems of conventional ESP fitted charges is that only point charges at the nuclei are employed, and this does not include the charge penetration effect. That is why only points outside the electron distribution of the molecules can be used for meaningful fitting. However, the charge penetration effect can be quite significant even outside the van der Waals surface, and various procedures have been proposed to include this effect in molecular modeling.[21-31] Including the charge penetration effect can greatly improve the description of the electrostatics for points within or close to the van der Waals surfaces of the molecules.

In the present study, we seek to derive ESP charges by using new screened charge models that include the charge penetration effect. Our new models are based on modified versions of the screened charge model proposed in a previous study.[29] The electron

densities around atoms are delocalized, and the atoms in the molecules are approximated as spherically symmetric delocalized charge distributions. In Section 8.2, we will introduce the screened charge models to include the charge penetration effect and describe the procedure to derive the ESP charges. Section 8.3 will give the computational details. Section 8.4 will present the results and a detailed analysis of the new methods. Section 8.5 will summarize the main conclusions.

8.2 Methods

8.2.1 Two kinds of screened charge models

We improve the point-charge model by considering the detailed electronic structure of the atoms. We make the approximation that the electrons in the atoms of the molecule, complex, or cluster under consideration are distributed as a sum of spherical distributions centered at the nuclei. Two kinds of screened models are proposed; we will call them outer-density screening (ODS) and full-density screening (FDS).

The ODS method was presented previously,[29] but it is reparametrized in the present work. In this model, we distribute a part of electron density in a Slater-type orbital (STO). At a distance r from its nucleus, the effective charge of atom A is written as:[29]

$$q_{A,\mathrm{eff}} = q_A + n_{\mathrm{screen}} f_{Z_A}(\zeta_{Z_A} r)\exp(-2\zeta_{Z_A} r) \tag{8.1}$$

where q_A is the partial atomic charge of atom A, ζ_{Z_A} is a parameter depending on the atomic number Z_A of atom A, n_{screen} is the number of electrons included in the STO, and f_{Z_A} is a polynomial factor derived in the previous paper.[29] The parameter n_{screen}

equals $1 - q$ for H atoms and 1 for other atoms.[29] The present model differs from the model proposed in the previous work in two aspects of the parametrization. First, metal atoms are screened in the present study, while they are not screened in the previous study. Second, ζ_{Z_A} values are recommended for all elements in the present study (see Sections 8.2.4 and 8.2.5).

One of the drawbacks of the ODS model is that only part of the penetration effect is included since only n_{screen} electrons are in the screening distribution. Therefore, in the second model, which is called the full-density screening (FDS) model, we approximate the charge density of an atom in a molecule, except for H atoms, as the sum of the charge density of the neutral atom and a charge distribution (which is positive for positive partial atomic charge and negative for negative partial atomic charge) in an STO. The analytical expressions for the effective charges of neutral atoms are taken from Strand and Bonham's work.[32] For an H atom, we assume that the atom is composed of a nucleus and an electron density distributed in the hydrogen 1s orbital, which is a special case of an STO.

The effective charge of a neutral atom A in the work of Strand and Bonham is expressed as:

$$Z_{\text{eff,A}}(r) = Z_A\left(\sum_i {}^{a}\gamma_i \exp(-{}^{a}\lambda_i r) + r\sum_j {}^{b}\gamma_i \exp(-{}^{b}\lambda_j r)\right) \tag{8.2}$$

where r is the distance from the nucleus, $Z_{\text{eff,A}}(r)$ is the effective charge of the neutral atom, Z_A is the atomic number, and γ_i and λ_i are fitted parameters. The partial atomic

charge q_A, which can be either positive (a hole distribution) or negative (a particle distribution), is distributed in an STO:

$$\varphi_{(n)} = Dr^{n-1}\exp(-\zeta_{Z_A} r) \tag{8.3}$$

where *n* is the period number of the element, and *D* is the factor that normalizes the orbital to unity. Thus, the effective charge due to the STO is

$$q_{A,\mathrm{STO}} = q_A - q_A f_{Z_A}(\zeta_{Z_A} r)\exp(-2\zeta_{Z_A} r) \tag{8.4}$$

where f_{Z_A} is the same polynomial factor as already introduced in eq 7.1.

Therefore, the total effective charge for all atoms except H atoms is obtained by adding eqs 2 and 4 to get

$$q_{A,\mathrm{eff}} = Z_{\mathrm{eff,A}}(r) + q_A - q_A f_{Z_A}(\zeta_{Z_A} r)\exp(-2\zeta_{Z_A} r) \tag{8.5}$$

For H atoms, the total effective charge can be derived from eq 7.1 by substituting $1 - q_A$ for n_{screen}:

$$q_{\mathrm{H,eff}} = q_{\mathrm{H}} + (1 - q_{\mathrm{H}}) f_{\mathrm{H}}(\zeta_{\mathrm{H}} r)\exp(-2\zeta_{\mathrm{H}} r) \tag{8.6}$$

Values of the exponents ζ_{Z_A} of the STOs will be recommended for all elements in Sections 8.2.4 and 8.2.5.

8.2.2 ESP charges

To derive the ESP charges from quantum mechanical (QM) electrostatic potentials, we define the error function to be minimized during the fitting process as

$$y(q_1, q_2, \ldots q_N) = \sum_{k=1}^{m} [V_k^{\mathrm{QM}} - V_k^{\mathrm{ESP}}(q_1, q_2, \ldots q_N)]^2 \tag{8.7}$$

where m is the total number of fitting points, N is the number of atoms in the molecule, V_k^{QM} is the QM electrostatic potential, q_i is the fitted partial atomic charge (i = 1, 2, …, N), and $V_k^{\mathrm{ESP}}(q_1, q_2, \ldots q_N)$ is the fitted electrostatic potential calculated from the point charges or the screened charges. The fitted electrostatic potential is expressed as

$$
\begin{aligned}
V_k^{\mathrm{ESP}}(q_1, q_2, \ldots q_N) &= \sum_{i=1}^{N} \frac{q_i}{r_{ik}} && \text{(point charge model)} \\
&= \sum_{i=1}^{N} \frac{q_i + n_{\mathrm{screen}} f(\zeta_i r_{ik}) \exp(-2\zeta_i r_{ik})}{r_{ik}} && \text{(ODS)} \\
&= \sum_{i=1,\ \text{not H atom}}^{N-N_{\mathrm{H}}} \frac{Z_{\mathrm{eff},i}(r_{ik}) + q_i - q_i f(\zeta_i r_{ik}) \exp(-2\zeta_i r_{ik})}{r_{ik}} \\
&\quad + \sum_{i=1,\ \text{H atom}}^{N_{\mathrm{H}}} \frac{q_i + (1 - q_i) f(\zeta_i r_{ik}) \exp(-2\zeta_i r_{ik})}{r_{ik}} && \text{(FDS)}
\end{aligned}
\tag{8.8}
$$

where N_{H} is the number of H atoms, r_{ik} is the distance between atom i and fitting point k, and ζ_i is the exponential parameter of the STO for atom i.

Finding the minimum of y subject to the conservation of charge is equivalent to finding the stationary points of the Lagrangian function z:

$$
z = y(q_1, q_2, \ldots q_N) + \lambda\left(\sum_{i=1}^{N} q_i - q_{\mathrm{tot}}\right) \tag{9}
$$

where q_{tot} is the total charge of the molecule. We solve the $n + 1$ linear equations obtained by $\partial z / \partial q_i = 0$ and $\partial z / \partial \lambda = 0$ to get the partial atomic charges.

8.2.3 Selection of fitting points

Another choice we need to make is which points to use for fitting the electrostatic potentials. We have tested 66 point selection schemes, and we will present results for six of them. The first two are the MK[10] and ChElPG[12] point selection schemes. They are well defined in the literature, and since a detailed description can be found in refs. 10 and 12, we only summarize them briefly here. In the MK scheme, the fitting points are selected using the Connolly surface algorithm from four shells of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radius.[10] In the ChElPG scheme, the fitting points are selected from a Cartesian grid around the molecule with nodes spaced 0.3 Å apart, including 2.8 Å of headspace on all sides and excluding the points that fall inside the van der Waals radius of any atom.[12] We will also briefly consider the earlier ChElP[11] scheme.

The other schemes tested here are formulated in various ways, and the four new schemes for which results are presented are based on selecting points from a Cartesian grid with nodes spaced by 0.2 Å. Different fitting points are selected from the grid of points in different schemes. In particular, we selected the fitting points k that satisfy:

$$\begin{aligned} \min_j \{r_{jk} - aR_j\} > 0 \\ \min_j \{r_{jk} - bR_j\} < 0 \end{aligned} \tag{8.10}$$

where $j = 1, 2, \ldots, N$, a and b are parameters that specify the scheme, r_{jk} is the distance between fitting point k and atom j, and R_j is the van der Waals radius of atom j, which is from the consistent van der Waals radii suggested in the *CRC Handbook of Chemistry and Physics*.[33] The distinguishing parameters of the four new schemes for which we

present results are given in Table 8.1. For example, in scheme C, the fitting points are in a shell with inner radius 1.4 R_i and outer radius 1.6 R_i around each atom *i*, but all points within 1.4 R_j of any other atom *j* are excluded; schemes are illustrated in Fig. 8.1.

8.2.4 Optimization of the point selection scheme and the ζ values of STOs for common elements

We optimize the ζ values for common elements such that the derived ESP charges can best reproduce the quantum mechanical electrostatic potentials. We started the optimization of the ζ values from the modified Strand and Bonham's parameters listed in Table 8.2, which are ζ = 1.32 for the H atom and half of the exponents for the outermost terms in ref. 32 for non-H atoms.[29,32] We multiplied the ζ parameters for all non-H atoms by a scaling factor *S*. Rather than optimizing the ζ value for each individual element, we optimize a common scaling factor for the best fit in the ODS and FDS models, respectively. The criterion for the best fit is explained next.

Because different point selection schemes use different fitting points, the value of the error function defined in eq 8.7 cannot be directly used as a criterion to compare the fitting performance among different point selection schemes. For testing, we used a Cartesian grid with nodes spaced by 0.3 Å. The testing points are grouped into layers around the molecule. Layer *L* includes all the points *k* that satisfy:

$$\begin{aligned} &\min_j\{r_{jk} - LR_j\} > 0 \\ &\min_j\{r_{jk} - (L+0.1)R_j\} < 0 \end{aligned} \tag{8.11}$$

for j = 1, 2, …, N. A sample layer is illustrated in Fig. 8.1.

Then, we define the relative root mean square error of electrostatic potentials in layer L in molecule j as

$$\varepsilon_{L,j} = \sqrt{\sum_{k=1}^{t} [V_k^{\mathrm{QM}} - V_k^{\mathrm{ESP}}(q_1, q_2, ...q_N)]^2 / \sum_{k=1}^{t} V_k^{\mathrm{QM}^2}} \tag{8.12}$$

where t is the number of testing points in layer L. Then we define the following averages of these errors. The average relative error over all 19 layers (L = 0.1, 0.2, …, 1.9) for molecule j is defined as

$$\varepsilon_{\mathrm{all},j} = \frac{1}{19} \sum_{L=0.1}^{1.9} \varepsilon_{L,j} \tag{8.13}$$

and the average relative error over the outer 10 layers (those beyond the van der Waals radius) for molecule j is defined as

$$\varepsilon_{\mathrm{outer},j} = \frac{1}{10} \sum_{L=1.0}^{1.9} \varepsilon_{L,j} \tag{8.14}$$

and the average relative errors over all molecules are

$$\varepsilon_{\mathrm{all}} = \frac{1}{M} \sum_{j=1}^{M} \varepsilon_{\mathrm{all},j} \tag{8.15}$$

$$\varepsilon_{\mathrm{outer}} = \frac{1}{M} \sum_{j=1}^{M} \varepsilon_{\mathrm{outer},j} \tag{8.16}$$

where M is the number of molecules considered (15 in the present study); $\varepsilon_{\mathrm{all}}$ and $\varepsilon_{\mathrm{outer}}$ show the performance of the fitting.

We tested 66 point selection schemes and various scaling factors, and the point selection schemes and optimized scaling factors (*S*) that give the smallest $\varepsilon_{\text{outer}}$ using the ODS and FDS models are shown in the last six rows of Table 8.3.

8.2.5 ζ values for other elements

The zeta values have units of a_0^{-1}, and they can be converted to effective radii in Å by

$$R_{\text{eff},Z} = C / \zeta_Z \tag{8.17}$$

where $C = 0.5292$ Å/a_0, and Z is atomic number.

For the ten elements in Table 8.2, we computed the average value of $R_{\text{vdW},Z} / R_{\text{eff},Z}$, where $R_{\text{vdW},Z}$ is the van der Waals radius (in Å) in the *CRC Handbook of Chemistry and Physics*,[33] we find that $< R_{\text{vdW},Z} / R_{\text{eff},Z} >= 2.93 \pm 0.30$, where the stated error is a standard deviation. Therefore, for other elements, we recommend that ζ_Z be set to 1.4 (the optimized scaling factor of the ODS method from Table 8.3) time 2.93 $R_{\text{vdW},Z}^{-1}$ (equals 4.10 $R_{\text{vdW},Z}^{-1}$) for the ODS method and that ζ_Z be set to 1.6 (the optimized scaling factor of the FDS method from Table 8.3) times 2.93 $R_{\text{vdW},Z}^{-1}$ (equals 4.69 $R_{\text{vdW},Z}^{-1}$) for the FDS method. Since $R_{\text{vdW},Z}$ is available for the whole periodic table, this defines ζ_Z for all elements.

8.3 Computational details

The electrostatic potentials at all the fitting points and the testing points are calculated using *Gaussian 09*,[34] and these potentials are used as input for electrostatic fitting with the screened charge models in a small code that solve the linear equations arising from eq 8.9. We have also modified link l602 of *Gaussian09* to make the output used by the fitting process more convenient.

For the molecules in the test suite, we used M06-2X[35,36]/6-31G*[37-39] optimized geometries with M06-2X/MG3S[40] single-point calculations. The test suite has 15 molecules, in particular $(CH_3)_2CO$, $(CH_3)_2SO$, CH_3CN, CH_3OH, H_2O, $HCONH_2$, NH_3, ClF, H_2S, $H_3SiOHAlH_3$, $H_3SiOSiH_3$, HCl, HF, PH_2CHCH_2, and PH_3.

8.4 Results and discussion

Three choices are required to define an ESP fitting method: (1) the charge model, which can be the point-charge (PC) method, the ODS method, or the FDS method; (2) the point selection scheme, which can be MK, ChElP, ChElPG or any of the 63 new schemes (for example schemes A–D of Table 8.1); and (3) the ζ values, which are determined by the scaling factor *S* for all non-H atoms. Therefore, we will use *X*-*Y*-*S* to specify a screened charge method, or *X*-*Y* for a point-charge method where no scaling factor is needed, where *X* represents the charge model, *Y* represents the point selection scheme, and *S* represents the scaling factor.

8.4.1 Overall fitting performance

Table 8.3 shows ε_{all}, $\varepsilon_{\text{outer}}$, and the mean unsigned errors (MUE) of their dipole moments for the point-charge model and two screened charge models. (The "error" in the dipole moment is defined as the difference between the dipole moment calculated from the electron density (as usual) and that calculated from the fitted charges.)

The first two rows of Table 8.3 show the results obtained by using the PC model with the MK and ChElPG point selection schemes. These are well-established and popular methods and are presented for comparison with the new methods.

The next two rows show the optimized point selection schemes and their optimized scaling factors that minimize $\varepsilon_{\text{outer}}$ for the ODS and FDS models, respectively. The optimum ODS scheme for $\varepsilon_{\text{outer}}$ is ODS-C-1.4, and the optimum FDS scheme for $\varepsilon_{\text{outer}}$ is FDS-C-1.6; however several other methods also give small $\varepsilon_{\text{outer}}$. The next four rows show the results with the MK and ChElPG point selection schemes using the ODS and FDS models, respectively. When we consider all these error criteria (ε_{all}, $\varepsilon_{\text{outer}}$, and MUE in dipole moments), the best scheme is FDS-MK-1.6. As compared to the point-charge method, the FDS method has about the same MUE (0.05 D) in dipole moments, but the error in the electrostatic potential averaged over the ten outer layers is reduced by a factor of 1.2 (0.27 vs. 0.32), and *the error in the electrostatic potential averaged over all 19 layers is reduced by a factor of 3* (0.21 vs. 0.62). The reason for this dramatic improvement is that the point-charge model becomes very unrealistic in regions close to a

nucleus. Since the FDS model includes penetration effects all the way in to the nucleus, it is more realistic than the ODS model in the inner regions.

Another key conclusion from Table 8.3 is that conventional point selection schemes originally developed for point-charge fitting, in particular MK and ChElPG, can also be used for electrostatic potential fitting in the screened charge models. Therefore conventional electrostatic fitting subroutines can be improved by simply replacing the point-charge model by the screened charge models. As these point selection schemes have been implemented in popular software, such as *Gaussian 09*,[34] minimum modifications are needed to use the screened charge methods. Our final recommendation is to us the FDS-MK-1.6 method to derive ESP charges.

8.4.2 Stability of the fitted charges

The sensitivity of the fitted charges to the positions of the fitting points is one of the most important criteria for comparing the quality of the point-charge model to that of the screened charge models. Since the charge penetration effect is included in the screened charge models, fitting electrostatic potentials to screened charges is more physical than using point charges, and therefore, the fitted charges can be more stable. As the MK, ChElP, and ChElPG point selection schemes use different points for fitting, we calculated the averaged standard deviations ($\overline{\sigma}$) of charges obtained by using these three point selection schemes, and we use these values as an indication for the sensitivity; the sensitivity $\overline{\sigma}$ for molecule *i* is defined as

$$\overline{\sigma}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\sqrt{\frac{1}{2}\left[(q_{i,j,\mathrm{MK}} - \overline{q}_{i,j})^2 + (q_{i,j,\mathrm{ChElP}} - \overline{q}_{i,j})^2 + (q_{i,j,\mathrm{ChElPG}} - \overline{q}_{i,j})^2\right]} \qquad (8.18)$$

where

$$\overline{q}_{i,j} = \frac{1}{3}(q_{i,j,\mathrm{MK}} + q_{i,j,\mathrm{ChElP}} + q_{i,j,\mathrm{ChElPG}}) \qquad (8.19)$$

where N_i is the number of atoms in molecule i, $q_{i,j,\mathrm{MK}}$, $q_{i,j,\mathrm{ChElP}}$ and $q_{i,j,\mathrm{ChElPG}}$ are the MK, ChElP, and ChElPG partial atomic charges (q_A) of atom j in molecule i, respectively, and $\overline{q}_{i,j}$ is the averaged MK, ChElP, and ChElPG partial atomic charge of atom j in molecule i.

The $\overline{\sigma}_i$ values for all molecules in the test suite are shown in Table 8.4. When $\overline{\sigma}_i$ is further averaged over 15 molecules, the value of 0.044 in the point-charge model is reduced to 0.024 in the ODS model and to 0.021 in the FDS model. This shows that *the charges obtained by fitting with the screened charge models are about a factor of 2 less sensitive to the positions of the fitting points than those obtained by the point-charge model.* This finding can also explain why the MK and ChElPG point selection schemes perform about equally as well as the optimized point selection scheme (scheme C) in the screened charge methods.

8.4.3 Case study ($HCONH_2$)

The formamide molecule, shown in Figure 8.2, is used as an example to demonstrate the performance of the screened charge models in detail. The effects of the fitting region and ζ values of STOs will be discussed. The PC model and two screened models (i.e.,

ODS and FDS) are tested. For the two screened charge models, a scaling factor of 1.4 is always used for the ODS model, and 1.6 is used for the FDS model, except in section 8.4.3.3 where we discuss the effect of the ζ values.

8.4.3.1 Electrostatic potential fitting using various methods

Figure 8.3 shows the relative root mean square error of electrostatic potentials ($\varepsilon_{L,\mathrm{HCONH2}}$) for each layer (see Fig. 8.2) when using the PC model and two screened charge models; $\varepsilon_{L,\mathrm{HCONH2}}$ is defined in eq 8.12. Results from the MK and ChElPG point selection schemes are shown. The PC model fails for the layers that are very close to the nuclei of atoms, and the screened charge methods improve the fitting over all layers. The ODS model improves the fitting in the layers around one van der Waals radius of the centers of atoms, but the error becomes large for the layers that are closer to the centers of atoms. The FDS model improves the fitting over all layers. It is pleasantly surprising that the FDS model can also describe the layers that are very close to the nuclei (even the L = 0.1 layer) very well.

Table 8.5 shows that there is no significant difference of partial atomic charges derived from the PC model and the screened charge models. In most cases, the advantage of the screened charge models for small molecules are more accurate electrostatic potentials, not more accurate partial atomic charges.

8.4.3.2 Effect of the fitting region

To further test how the fitting region affects the fitting performance of the electrostatic potentials and the fitted charges and dipole moment in the screened charge models, we consider a series of point selection schemes by varying the fitting region (see Table 8.1). Only the FDS model is tested in this section. Figure 8.4 shows the relative root mean square error of electrostatic potentials for each layer ($\varepsilon_{L,\mathrm{HCONH2}}$) in the four schemes, and Table 8.6 shows the fitted charges and dipole moments in the four schemes. Figure 8.4 shows that the largest $\varepsilon_{L,\mathrm{HCONH2}}$ occurs in layers 0.6, 0.7 and 0.8 in all schemes, possibly because the charge distributions of atoms in the molecule change significantly in the bonding region, and therefore it is harder to describe these regions using the present model. When the fitting points are very close to the atoms, such as in scheme A, the electrostatic potentials of the outer layers cannot be well fitted, and the fitted dipole moments do not agree with the directly calculated ones. When the fitting points are outside one van der Waals radius of the nuclei of any atom, such as in schemes B–D, the fitting performance is good for all layers, and the fitted dipole moments are close to the directly calculated ones. Therefore, only the points outside the van der Waals surface are suitable for electrostatic potential fitting in the screened charge models, and the ChElPG and MK point selection schemes satisfy this criterion.

8.4.3.3 Effect of the ζ values of STOs

An important parameter we used in the fitting is the scaling factor *S*. The scaling factor determines the ζ values of the STOs for all non-H atoms. We tested the fitting

performance using various scaling factors. The MK point selection scheme is used for all cases in these figures. Figures 8.6 and 8.7 show how the scaling factor affects the relative root mean square error of electrostatic potentials ($\varepsilon_{L,\mathrm{HCONH2}}$) for each layer in the ODS and FDS models, respectively. Compared with the optimized scaling factor, a scaling factor of 100 or 1 gives similar results for the outer layers but much worse results for the inner layers in both the ODS and FDS models. Therefore, a reasonable set of ζ values of STOs in the charge model is necessary for a good fit of electrostatic potentials in all the physical regions. Table 8.7 shows how the scaling factor affects the fitted charges and dipole moment in the ODS and FDS models. A small scaling factor (i.e., 1) can lead to inaccurate fitted charges and dipole moments, and it is safer to use our recommended scaling factor to derive physical charges.

In the previous study,[29] the ODS model with a scaling factor of 1 for all non-metal elements showed good performance for the electrostatic interactions. However, since the charge penetration effect is not fully included in the ODS model, the ODS model with a scaling factor of 1 does not fit the electrostatic potentials well. This will be discussed in more detail in section 8.4.6.

8.4.4 Restraints on non-polar H atoms

Although in many cases the charges derived from the point-charge model and the screened charge models are similar, we found that in some cases the charges from the point-charge model and screened charge models are different. An example is $(CH_3)_2CO$, shown in Figure 8.7. Table 8.8 shows the fitted charges and dipole moments. Comparing

the results using the MK and ChElPG point selection schemes, we found that though the fitted charges and dipole moments derived from the screened charge models are more stable than those from the point-charge model, there are some systematic differences between the results from the point-charge model and the screened charge models. In particular, the C–H bonds are more polar in the screened charge models than those in the point-charge model. In previous studies, restraints have been added on certain atoms to get more reliable charges, such as in the RESP method.[13,14] In the present study, we explore the use of restraints on the non-polar H atoms.

We modified the error function in eq 8.7 by adding the restraints:

$$y(q_1, q_2, \ldots q_N) = \sum_{k=1}^{m} [V_k^{\mathrm{QM}} - V_k^{\mathrm{ESP}}(q_1, q_2, \ldots q_N)]^2 + \lambda' \sum_{\substack{i=1 \\ \text{non-polar H atom}}}^{N_{\mathrm{NPH}}} q_i^2 \qquad (8.20)$$

where N_{NPH} is the number of non-polar H atoms.

We tested the restraints for two well performing schemes, which are the ODS model with a scaling factor of 1.4 and the FDS model with a scaling factor of 1.6, in both cases using the MK point selection scheme. Table 8.9 shows the fitted charges, dipole moments, and fitting performance with various restraints on non-polar H atoms. It is found that we can significantly decrease the charges on H atoms compared with the results from the point-charge model while maintaining small $\varepsilon_{\mathrm{all},(\mathrm{CH_3})_2\mathrm{CO}}$ and $\varepsilon_{\mathrm{outer},(\mathrm{CH_3})_2\mathrm{CO}}$. A restraint of $\lambda' = 0.01$ in the ODS and FDS models is suggested for non-polar H atoms.

8.5 Applications to charged molecules and molecules containing s-block and d-block elements

The tests described so far are for neutral molecules containing H atoms and p-block elements. To further test the performance of the proposed screened charge schemes, we studied several charged molecules and molecules containing s-block and d-block elements. The geometries and electrostatic potentials are from M06[35,36]/def2-TZVP[41] calculations. The fitted partial atomic charges and fitting performances are shown in the Supporting Information, and the conclusions drawn from these studies are summarized in the rest of this section.

For the three charged molecules, including ClO_3^-, H_3O^+, NO_3^-, the two screened charge schemes give more stable charges and better fit to the electrostatic potentials than the point-charge model. We conclude that the present method can be applied to the charged species in the same way as it is applied to neutral ones.

For the test of a set of molecules containing s-block and d-block elements, we found that the performance is not as good as for the H atom and p-block elements; however the method still appears useful. In addition to the FDS method, the tables show the results for the ODS method and also for a method denoted as "ODS (H and p-block)" in the tables—the last named option corresponds to the ODS scheme with only the H atom and p-block atoms being screened; for 14 test cases, with both MK and ChElPG point selection, that makes a total of 84 possible comparisons to the point charge method, and in all 84 cases, we find that ε_{all} is improved. However, when one looks only at $\varepsilon_{\text{outer}}$, we find that FDS

improves the results only about half the time, ODS improves them most of the time, and ODS (H and p-block) always improves them or at least does as well as point charges. Therefore, on average, for molecules containing s-block and d-block elements, the ODS model works better than the FDS model and the point-charge model. We think that for s-block and d-block elements, the charge penetration effects are more difficult to take into account, and certainly the present models oversimplify the problem. The present models should be used with caution for systems containing s and d block elements, although the ODS (H and p-block) scheme seems safe, and FDS appears safe when the s-block and d-block elements are surrounded by H and p-block elements.

8.6 Limitations and improvements and comments on combined QM/MM calculations

Although the two screened charge models improve the fitting performance of the electrostatic potentials, both models have disadvantages.

In the ODS model, the charge density of an atom is assumed to be the sum of a point-charge and a smeared charge. This model has been shown to be a good compromise for calculating the QM/MM electrostatic energy and avoiding overpolarization in the QM/MM method.[29] However, this model only includes part of the charge penetration effect since only n_{screen} electrons are in the screening distribution. Since the previous ζ values of STOs were parametrized to best reproduce the directly calculated electrostatic and induction energies,[29] these parameters compensate for the approximation of the ODS model. Therefore, the same set of parameters does not reproduce the electrostatic

potentials as well as the new parameters optimized for that purpose in the present work, as shown in section 8.4.3.3. To test how well the new set of parameters works for the QM/MM electrostatic and induction energies, we carried out the same QM/MM calculations as presented in ref. 29 but now with the new parameters. The comparison is shown in Tables 8.10 and 8.11; as in ref. 29, we separate the tests into equilibrium geometries, compressed geometries (the severest test), and expanded geometries; and we show two results for two different basis sets for the QM region: acTZ, which is short for aug-cc-pVTZ,[42,43] and def2-TZVP.[41] The last column is averaged over the previous twelve columns and provides a summary. The first three rows of these two tables are from Tables 8.6 and 8.7 of ref. 29, and the last rows of these two tables show the results with the new parameters (i.e., 1.32 for the H atom and a scaling factor of 1.4 for non-H atoms). The new set of parameters performs worse than the optimized parameters for the QM/MM electrostatic energy and induction energy, but still greatly improves the results compared with the point-charge method. Therefore, the ODS model with the new set of parameters can be used to calculate both the QM/MM electrostatic and induction energies and the electrostatic potentials, while the previous set of parameters in ref. 29 should only be used to calculate the QM/MM electrostatic and induction energies.

In the FDS model, the delocalization character of the electrons is well described, even for the regions that are close to the nuclei of atoms in a molecule. The fitted charges can be useful to calculate the electrostatic energies in MM calculations. However, the FDS screened charge model may not be useful for QM/MM calculations if no Pauli exchange is added in SCF calculations because it greatly overestimates the magnitudes of induction

energies when one includes induction without exchange repulsion. This is more of a comment on the incompleteness of the combined QM/MM method, as usually implemented, than a limitation of the FDS screened electrostatic model.

Another issue to be considered with regard to the FDS model is that the charge density of the neutral atom may not be a good reference density for highly charged atoms. The charge density of a highly charged atom can be significantly different from that of the neutral atom, and the current model may not be appropriate. One way to overcome this drawback of the FDS model is to take the effective charge as the linear combination of the neutral atom and the charged atoms, with the weighting factor determined by the charge q_A. The effective charge of atom A is expressed as:

$$\begin{aligned} q_A^*(r) = \theta(-q_A)\left[(-q_A)Z_{\text{eff}}^{\text{anion}}(r) + (1+q_A)Z_{\text{eff}}^{\text{neutral}}(r)\right] \\ + \theta(q_A)\left[(q_A)Z_{\text{eff}}^{\text{cation}}(r) + (1-q_A)Z_{\text{eff}}^{\text{neutral}}(r)\right] \end{aligned} \tag{8.21}$$

where $Z_{\text{eff}}^{\text{anion}}(r)$, $Z_{\text{eff}}^{\text{neutral}}(r)$, $Z_{\text{eff}}^{\text{cation}}(r)$ are the effective charges of the atom in its anion, neutral, and cation forms, and $\theta(x)$ denotes a Heaviside function of x.

8.7 Conclusions

In the present study, we propose a new kind of screened charge model to derive partial atomic charges using ESP fitting, and we compare its performance to the conventional point-charge model and to a previous screened charge method. In the screened charge models, the charges are more stable with respect to the point selection schemes, and the fitting performance is better than with the point-charge model.

The ODS model with a scaling factor (*S*) of 1.4 and the FDS model with a scaling factor of 1.6, combined with either the ChElPG or MK point selection scheme, have shown good fitting performance and are recommended for use. In particular, our final recommended method for electrostatic potentials is FDS-MK-1.6. However, the scheme should be used carefully if there are s-block and d-block elements in the molecules. In those cases, ODS-MK-1.4 gives better performance based on our experience. Our recommended method for combined QM/MM calculations is ODS-ChElPG-*S* with $1.0 \le S \le 1.4$.

8.8 References for Chapter 8

1. Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87–110.
2. Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358–3363.
3. Mulliken, R. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
4. Baker, J. *Theor. Chem. Acc.* **1985**, *68*, 221–229.
5. Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129–138.
6. Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbo-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.
7. Lillestolen, T. C.; Wheatley, R. J. *Chem. Comm.* **2008**, 5909–5911.
8. Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592–601.
9. Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304–323.
10. Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
11. Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894–905.
12. Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.
13. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
14. Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 7300–7307.
15. Hu, H.; Lu, Z. Y.; Yang, W. *J. Chem. Theory Comput.* **2007**, *3*, 1004–1013.
16. Chen, D. L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 10225–10233.
17. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133–151.
18. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.

19. Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046–2054.
20. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367–383.
21. Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968–1986.
22. Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300–7306.
23. Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353–10359.
24. Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Phys. Chem. B* **2006**, *110*, 13682–13684.
25. Piquemal, J. P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.
26. Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. *J. Phys. Chem. A* **2007**, *112*, 268–280.
27. Cisneros, G. A.; Tholander, S. N.-I.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J. P. *Int. J. Quantum Chem.* **2008**, *108*, 1905–1912.
28. Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190–202.
29. Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330–3342.
30. Stone, A. J. *J. Phys. Chem. A* **2011**, *115*, 7017–7027.
31. Tafipolsky, M.; Engels, B. *J. Chem. Theory Comput.* **2011**, *7*, 1791–1803.
32. Strand, T. G.; Bonham, R. A. *J. Chem. Phys.* **1964**, *40*, 1686–1691.
33. Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. Atomic Radii of the Elements. In *CRC Handbook of Chemistry and Physics*; 92nd; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2011-2012; pp. 9-49f.
34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09, version A. 02*; Gaussian, Inc.: Wallingford CT, 2009.
35. Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157–167.

36. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
37. Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
38. Dill, J. D.; Pople, J. A. *J. Chem. Phys.* **1975**, *62*, 2921–2923.
39. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
40. Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384–1388.
41. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
42. Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
43. Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

Table 8.1 Four new schemes in the present study

scheme	parameters	
	A	*b*
A	0.6	0.8
B	1.0	1.2
C	1.4	1.6
D	1.8	2.0

Table 8.2 Modified Strand Bonham parameters (in a_0^{-1})[a]

atom	H	C	N	O	F
ζ	1.32	0.87	1.01	1.12	1.24
atom	Al	Si	P	S	Cl
ζ	0.68	0.74	0.81	0.88	0.95

[a] The optimized exponent for the H atom in Ref. 29 and half of the exponents for the outermost orbitals of Ref. 32 for non-H ato

Table 8.3 ε_{all} and $\varepsilon_{\text{outer}}$ and the mean unsigned error (MUE) of dipole moments (debyes) over 15 molecules using the point-charge model and the ODS and FDS models

charge model	point selection scheme	*S*	ε_{all}	$\varepsilon_{\text{outer}}$	MUE of dipole moment
PC[a]	MK	–	0.62	0.32	0.05
PC	ChElPG	–	0.61	0.32	0.07
ODS	scheme C	1.4	0.42	0.27	0.06
FDS	scheme C	1.6	0.21	0.27	0.06
ODS	MK	1.4	0.42	0.27	0.06
ODS	ChElPG[b]	1.4	0.41	0.27	0.07
FDS[b]	MK[b]	1.6	0.21	0.27	0.05
FDS	ChElPG	1.6	0.21	0.27	0.06

[a] Point-charge method [b] These are the finally recommended methods.

Table 8.4 Averaged standard deviation ($\overline{\sigma}_i$) of charges derived from the MK, ChElP, and ChElPG point selection schemes using the point-charge model and the ODS and FDS models

charge model	PC	ODS[a]	FDS[b]
$(CH_3)_2CO$	0.044	0.028	0.012
$(CH_3)_2SO$	0.174	0.113	0.073
CH_3CN	0.051	0.031	0.038
CH_3OH	0.028	0.005	0.015
ClF	0.014	0.010	0.010
H_2O	0.002	0.002	0.003
H_2S	0.023	0.007	0.018
$H_3SiOHAlH_3$	0.081	0.034	0.033
$H_3SiOSiH_3$	0.072	0.035	0.056
HCl	0.012	0.012	0.012
$HCONH_2$	0.053	0.030	0.011
HF	0.010	0.009	0.008
NH_3	0.012	0.009	0.004
PH_2CHCH_2	0.061	0.027	0.012
PH_3	0.018	0.006	0.004
Averaged over 15 molecules	0.044	0.024	0.021

[a] The ODS model with a scaling factor of 1.4

[b] The FDS model with a scaling factor of 1.6

Table 8.5 Partial atomic charges and dipole moments of $HCONH_2$ derived from the point-charge model and the ODS and FDS models. The MK and ChElPG point selection schemes are tested. The directly calculated dipole moment is 4.01 Debye.

charge model	PC		ODS		FDS	
point selection scheme	MK	ChElPG	MK	ChElPG[a]	MK[a]	ChElPG
scaling factor *S*	–	–	1.4	1.4	1.6	1.6
C1	0.61	0.67	0.60	0.62	0.59	0.59
H2	0.02	–0.01	0.02	0.01	0.02	0.02
O3	–0.54	–0.56	–0.53	–0.54	–0.53	–0.53
N4	–0.91	–0.89	–0.91	–0.88	–0.91	–0.88
H5	0.43	0.41	0.43	0.41	0.43	0.42
H6	0.39	0.38	0.39	0.38	0.39	0.38
Dipole moment (Debye)	3.98	3.99	3.98	3.98	3.99	3.99

[a] Recommended methods

Table 8.6 Partial atomic charges and dipole moments of $HCONH_2$ derived from various fitting regions. The directly calculated dipole moment is 4.01 Debye.

scheme	A	B	C	D
C1	0.43	0.57	0.58	0.56
H2	0.03	0.02	0.03	0.04
O3	–0.46	–0.52	–0.53	–0.52
N4	–0.68	–0.85	–0.91	–0.89
H5	0.36	0.42	0.43	0.43
H6	0.32	0.37	0.39	0.39
Dipole moment	3.75	3.97	4.00	4.00

Table 8.7 Partial atomic charges and dipole moments of HCONH2 derived from the point-charge model and the ODS and FDS models using various scaling factors. The MK point selection scheme is used for all cases. The directly calculated dipole moment is 4.01 Debye.

charge model	PC	ODS		FDS	
scaling factor *S*	–	1	100	1	100
C1	0.61	0.51	0.61	0.67	0.59
H2	0.02	0.06	0.02	0.01	0.02
O3	–0.54	–0.50	–0.54	–0.56	–0.53
N4	–0.91	–0.93	–0.91	–0.98	–0.91
H5	0.43	0.45	0.43	0.45	0.43
H6	0.39	0.41	0.39	0.40	0.39
Dipole moment (Debye)	3.98	4.04	3.97	3.96	3.99

Table 8.8 Partial atomic charges and dipole moments of $(CH_3)_2CO$ derived from the point-charge model and the ODS and FDS models using the MK and ChElPG point selection schemes. The directly calculated dipole moment is 3.10 Debye.

Charge model	PC		ODS		FDS	
point selection scheme	MK	ChElPG	MK	ChElPG	MK	ChElPG
scaling factor *S*	–	–	1.4	1.4	1.6	1.6
C1	0.75	0.70	0.78	0.72	0.80	0.76
O2	–0.54	–0.54	–0.54	–0.53	–0.54	–0.52
C3	–0.55	–0.40	–0.61	–0.53	–0.67	–0.65
H4	0.15	0.10	0.16	0.14	0.18	0.18
H5	0.15	0.10	0.16	0.14	0.18	0.18
H6	0.16	0.12	0.17	0.15	0.19	0.18
C7	–0.56	–0.40	–0.62	–0.53	–0.69	–0.65
H8	0.15	0.10	0.16	0.14	0.18	0.18
H9	0.15	0.10	0.16	0.14	0.18	0.18
H10	0.16	0.12	0.17	0.15	0.19	0.18
Dipole moment (Debye)	3.11	3.09	3.10	3.10	3.10	3.11

Table 8.9 Charges, dipole moments, ε_{all}, and ε_{outer} of $(CH_3)_2CO$ derived with various restraints on non-polar H atoms. The directly calculated dipole moment is 3.10 Debye.

Charge model	PC	ODS				FDS			
λ'	0	0	0.001	0.005	0.01	0	0.001	0.005	0.01
C1	0.75	0.78	0.76	0.70	0.65	0.80	0.78	0.72	0.66
O2	−0.54	−0.54	−0.54	−0.53	−0.52	−0.54	−0.54	−0.53	−0.52
C3	−0.55	−0.61	−0.58	−0.48	−0.39	−0.67	−0.64	−0.52	−0.42
H4	0.15	0.16	0.15	0.13	0.10	0.18	0.17	0.14	0.11
H5	0.15	0.16	0.15	0.13	0.10	0.18	0.17	0.14	0.11
H6	0.16	0.17	0.16	0.14	0.12	0.19	0.18	0.15	0.13
C7	−0.56	−0.62	−0.59	−0.48	−0.39	−0.69	−0.65	−0.53	−0.43
H8	0.15	0.16	0.15	0.13	0.10	0.18	0.17	0.14	0.11
H9	0.15	0.16	0.15	0.13	0.10	0.18	0.17	0.14	0.11
H10	0.16	0.17	0.17	0.14	0.12	0.19	0.18	0.15	0.13
Dipole moment (Debye)	3.11	3.10	3.10	3.09	3.08	3.10	3.10	3.09	3.09
ε_{all}	0.47	0.27	0.26	0.26	0.26	0.10	0.10	0.12	0.14
ε_{outer}	0.09	0.06	0.06	0.07	0.08	0.06	0.06	0.07	0.08

Table 8.10 Mean signed error (MSE) and mean unsigned error (MUE) of electrostatic energies (kcal/mol) using the QM/MM method[a]

	equilibrium /acTZ		equilibrium /def2-TZVP		compressed /acTZ		compressed /def2-TZVP		extended /acTZ		extended /def2-TZVP		All	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
PC[b]	6.43	6.43	6.43	6.43	14.80	14.80	15.10	15.10	3.02	3.02	2.98	2.98	8.13	8.13
Opt[c]	1.58	1.97	1.32	2.17	3.64	5.09	2.61	5.31	1.06	1.11	1.03	1.16	1.84	2.81
MSB[d]	1.59	2.22	1.19	2.29	3.55	5.39	2.15	5.90	1.07	1.25	1.00	1.24	1.71	3.06
ODS-1.4[e]	4.07	4.08	3.85	4.00	8.15	8.42	7.79	8.33	2.20	2.20	2.15	2.15	4.68	4.86

[a] Accurate values are SAPT electrostatic energies. The first three rows are from Table 6 of ref. 29, and the last row shows the results calculated in the present study.

[b] point-charge scheme

[c] ODS with optimized parameters of ref. 29

[d] ODS with modified Strand-Bonham (MSB) parameters

[e] ODS with a scaling factor of 1.4

Table 8.11 MSE and MUE of the induction energies (kcal/mol) using the QM/MM method[a]

	equilibrium /acTZ		equilibrium /def2-TZVP		compressed /acTZ		compressed /def2-TZVP		extended /acTZ		extended /def2-TZVP		All	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
PC[b]	0.67	1.09	1.12	1.18	3.46	3.98	4.34	4.46	0.12	0.45	0.35	0.38	1.72	1.94
Opt[c]	–0.74	1.00	0.69	0.82	–0.62	2.67	2.68	3.13	–0.34	0.45	0.25	0.30	0.46	1.40
MSB[d]	–0.79	1.02	0.68	0.80	–0.71	2.65	2.60	3.12	–0.36	0.46	0.25	0.29	0.42	1.39
ODS-1.4[e]	–0.28	1.01	0.75	0.82	0.77	2.56	3.14	3.31	–0.20	0.48	0.24	0.28	0.84	1.42

[a] Accurate values are SAPT damped induction energies. The first three rows are from Table 7 of ref. 29, and the last row shows the results calculated in the present study.

[b] point-charge scheme

[c] ODS with optimized parameters of ref. 29

[d] ODS with modified Strand-Bonham (MSB) parameters

[e] ODS with a scaling factor of 1.4

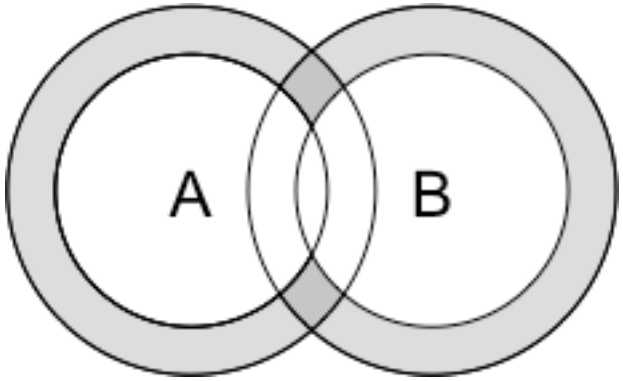


Figure 8.1 For schemes A–D of Table 8.1, we include points in the shaded area, where the inner sphere around each atom *j* has radius aR_j, and the outer sphere has radius bR_j. This shaded area also serves to illustrate the testing layers of eq 8.11. Layer L has inner sphere radius LR_j and outer sphere radius $(L + 0.1)\ R_j$.

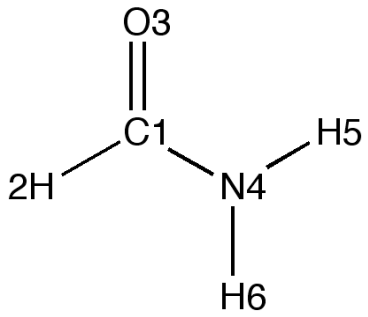


Figure 8.2 Atom numbering in formamid

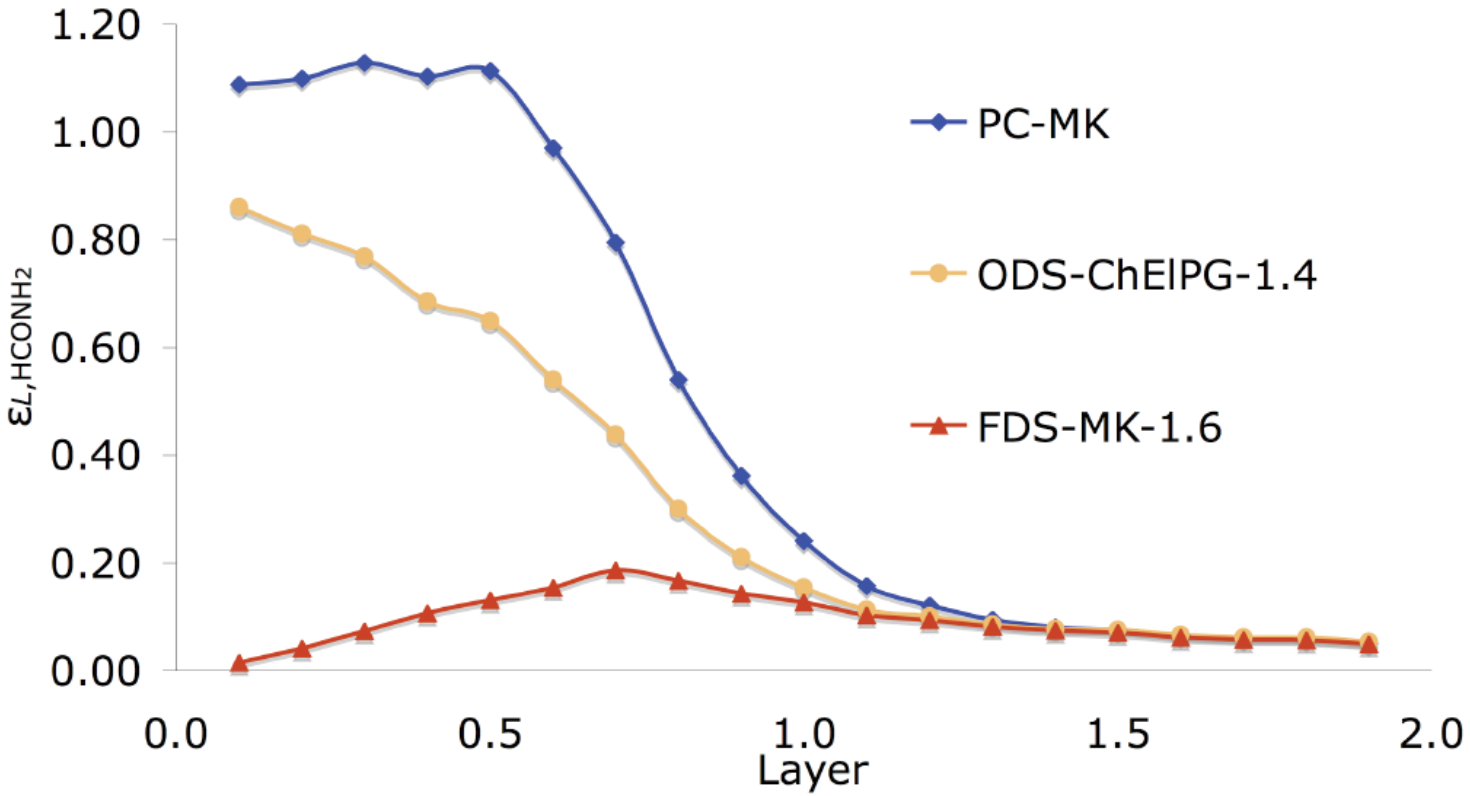


Figure 8.3 Relative root mean square error of electrostatic potentials for each layer using the MK and ChElPG point selection schemes with the point-charge model and the ODS and FDS models. The PC-ChElPG, ODS-MK-1.4, and FDS-ChElPG-1.6 methods (not shown) yield results almost superimposable on the three curves shown.

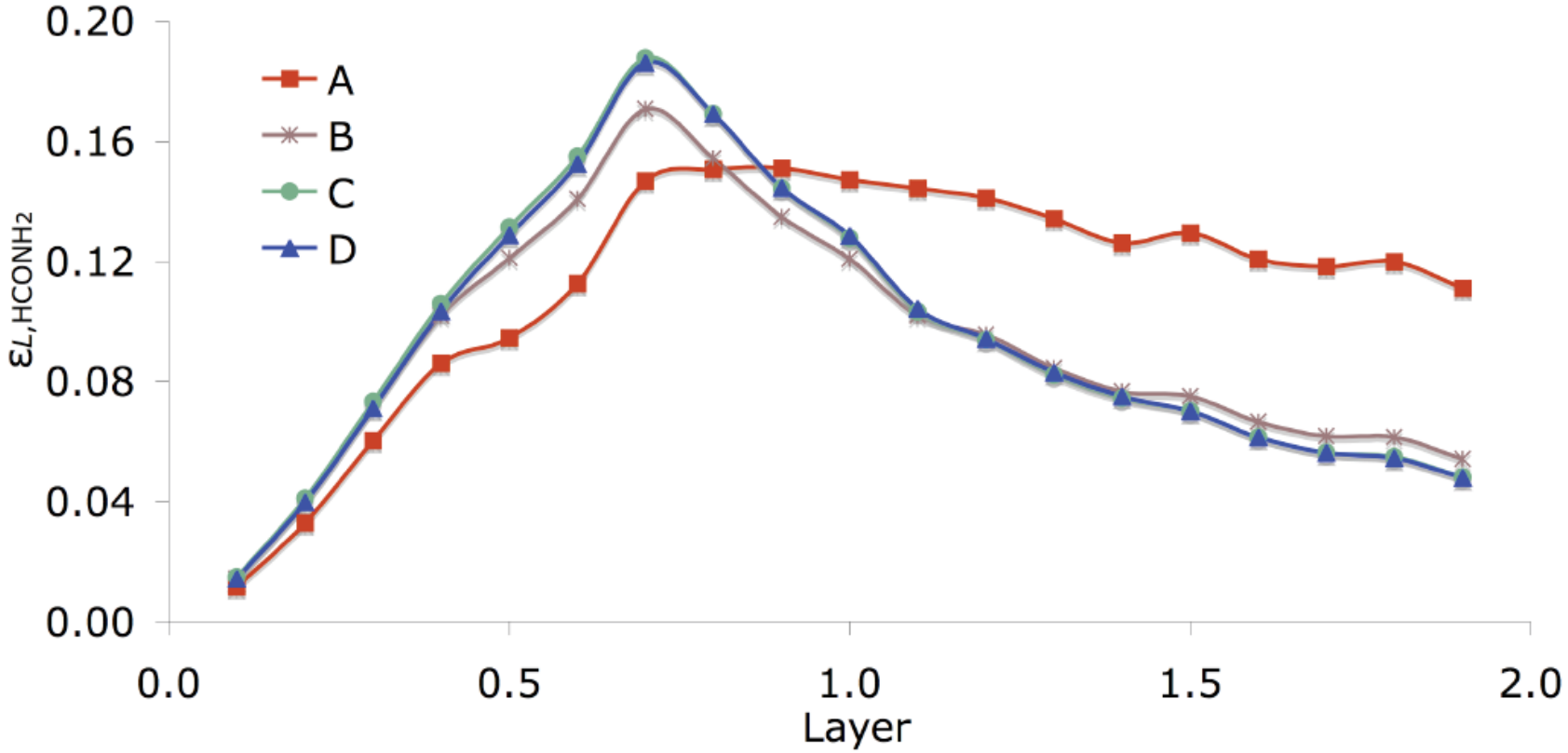


Figure 8.4 Relative root mean square error of electrostatic potentials by the FDS method for each layer using various fitting regions

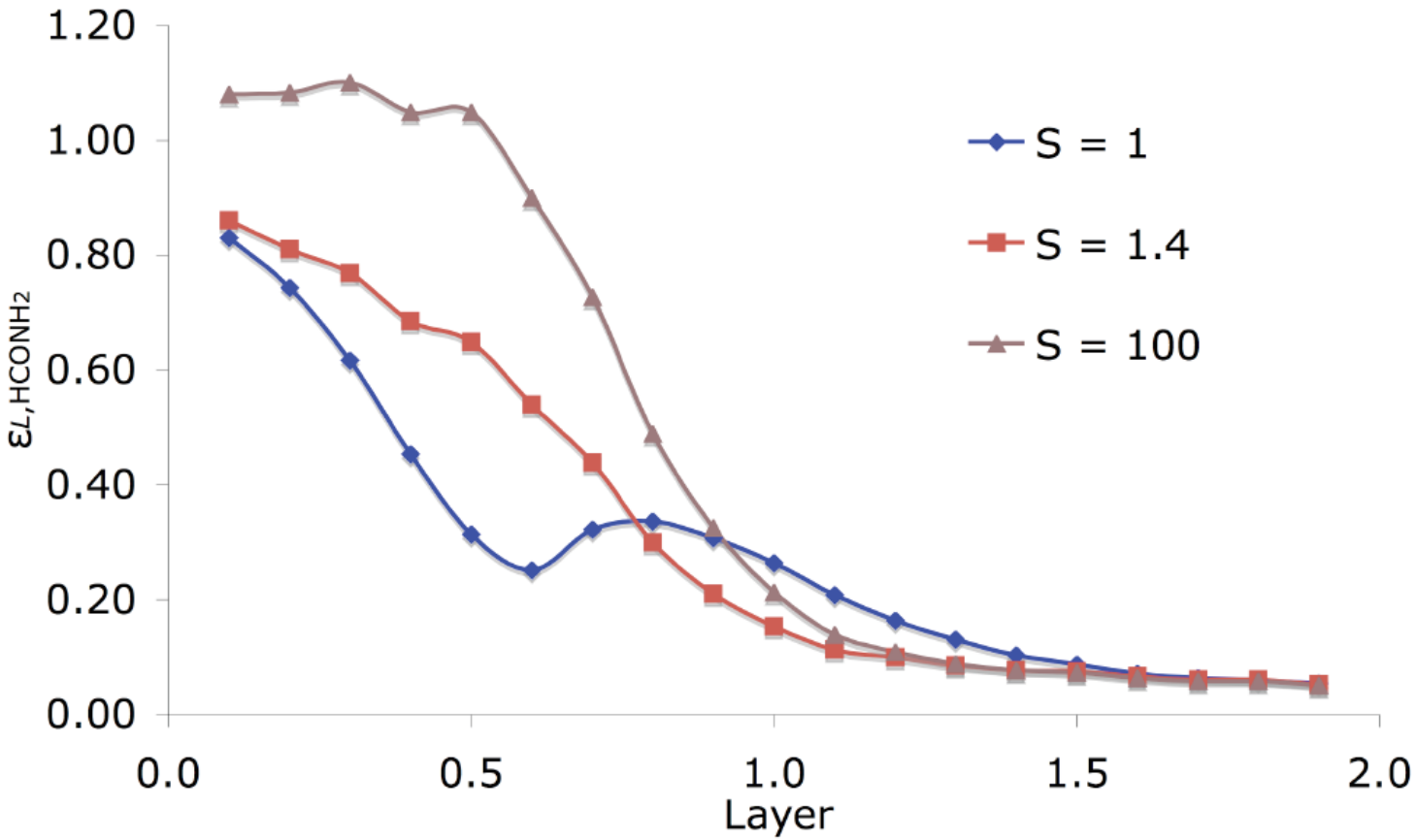


Figure 8.5 Relative root mean square error of electrostatic potentials over each layer using various scaling factors *S* in the ODS model

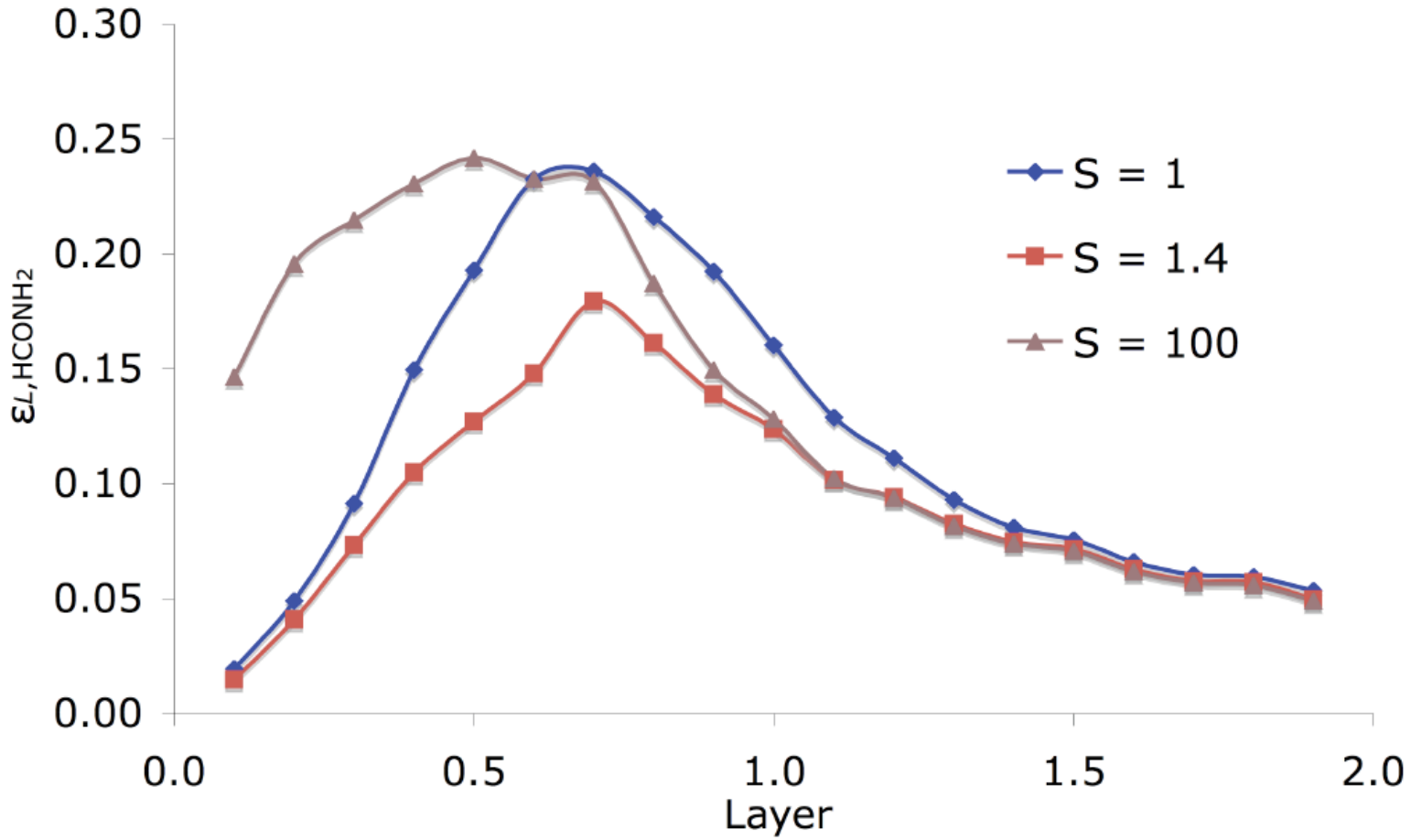


Figure 8.6 Relative root mean square error of electrostatic potentials over each layer using various scaling factors *S* in the FDS model

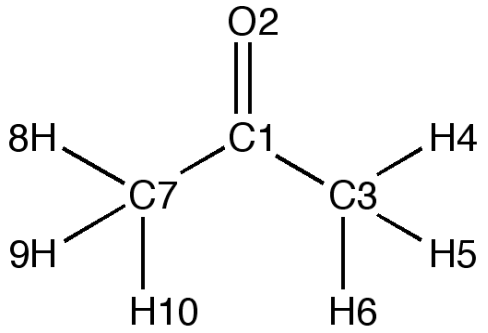


Figure 8.7 Atom numbering in acetone

Bibliography

Alary, F.; Poteau, R.; Heully, J. L.; Barthelat, J. C.; Daudey, J. P. *Theor. Chem. Acc.* **2000**, *104*, 174.
Alexander, D.; Mackerell, J. *J. Comp. Chem.* **2004**, *25*, 1584.
Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.
Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
Archambault, F.; Chipot, C.; Soteras, I.; Luque, F. J.; Schulten, K.; Dehez, F. *J. Chem. Theory Comput.* **2009**, *5*, 3022.
Assfeld, X.; Rivail, J.L. *Chem. Phys. Lett.* **1996**, *263*, 100.
Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192.
Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.
Baker, J. *Theor. Chem. Acc.* **1985**, *68*, 221.
Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1988**, *89*, 5739.
Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
Beran, G. *J. Chem. Phys.* **2009**, *130*, 164115.
Bernstein, N.; Kermode, J. R.; Csányi, G. *Rep. Prog. Phys.* **2009**, *72*, 026501.
Besler, B. H.; Merz Jr. K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *1*, 431.
Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
Boekfa, B.; Choomwattana, S.; Khongpracha, P.; Limtrakul, J. *Langmuir* **2009**, *25*, 12990.
Bredow, T.; Geudtner, G.; Jug, K. *J. Chem. Phys.* **1996**, *105*, 6395.
Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbo-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.
Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2010**, *6*, 2469.
Chen, D. L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 10225.
Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.
Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.
Cinquini, F.; Valentin, C. D.; Finazzi, E.; Giordano, L.; Pacchioni, G. *Theor. Chem. Acc.* **2007**, *117*, 827.
Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Phys. Chem. B* **2006**, *110*, 13682.
Cisneros, G. A.; Tholander, S. N.-I.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J. P. *Int. J. Quantum Chem.* **2008**, *108*, 1905.
Clementi, E.; Raimondi, D. L. *J. Chem. Phys.* **1963**, *38*, 2686.
Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Röhrig, U.; Spiegel, K.; Sulpizi, M.; VandeVondele, J.; Zumstein, M.; Röthlisberger, U. *Chimia* **2002**, *56*, 13.
Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.
Cramer, C. J. Essentials of Computational Chemistry: Theories and Models, 2nd ed.; John Wiley & Sons Inc, **2004**.
Curtiss, L. A.; McGrath, M. P.; Blaudeau, J. P.; Davis, N. E.; Binning, R. C.; Radom, L. *J. Chem. Phys.* **1995**, *103*, 6104.
Cusachs, L. C.; Trus, B. L.; Carroll, D. G.; Mcglynn, S. P. Int. *J. Quantum Chem. Symp.* **1967**, *1*, 423.
Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 10595.
Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
Dahlke, E. E.; Orthmeyer, M. A.; Truhlar, D. G. *J. Phys. Chem. B* **2008**, *112*, 2372.
Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1.
Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.
Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. Theochem.* **1999**, *461*, 1.
Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.
Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, *112*, 2063.
Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
DiLabio, G. A.; Wolkow, R. A.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 044708.
Dill, J. D.; Pople, J. A. *J. Chem. Phys.* **1975**, *62*, 2921.
Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.
Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
Elango, M.; Subramanian, V.; Rahalkar, A. P.; Gadre, S. R.; Sathyamurthy, N. *J. Phys. Chem. A* **2008**, *112*, 7699.
Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190.
Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.
Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2007**, *28*, 1476.
Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A.* **2007**, *111*, 6904.
Fedorov, D. G.; Jensen, J. H.; Deka, R. C.; Kitaura, K. *J. Phys. Chem. A* **2008**, *112*, 11808.
Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.
Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.
Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367.

Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300.
Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165.
Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
Gao, J. *Rev. Comp. Chem* **1996**, *7*, 119.
Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.
Gao, J.; Thompson, M. A., Eds.; Combined Quantum Mechanical and Molecular Mechanical Methods; ACS Symposium Series 712; American Chemical Society: Washington, DC, **1998**.
Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
Gill, P. M. W. *J. Phys. Chem.* **1996**, *100*, 15421.
Gomes, J. R. B.; Illas, F.; Hernández, N. C.; Márquez, A.; Sanz, J. F. *Phys. Rev. B* **2002**, *65*, 125414.
Gomes, J. R. B.; Illas, F.; Hernández, N. C.; Sanz, J. F.; Wander, A.; Harrison, N. M. *J. Chem. Phys.* **2002**, *116*, 1684.
Gomes, J. R. B.; Lodziana, Z.; Illas, F. *J. Phys. Chem. B* **2003**, *107*, 6411.
Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. *J. Phys. Chem. B* **2009**, *113*, 9646.
Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632.
Guillot, B.; Guissani, Y. *J. Chem. Phys.* **2001**, *114*, 6720.
Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
Hall, G. G.; Smith, C. M. *Int. J. Quantum Chem.* **1992**, *42*, 1237.
Hammes-Schiffer, S. *Curr. Opin. Struct. Biol.* **2004**, *14*, 192.
Hancock, G. C.; Truhlar, D. G.; Dykstra, C. E. *J. Chem. Phys.* **1988**, *88*, 1786.
Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285.
Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
Hehre, W. J.; Ditchfield R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231.
Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 790.
Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2925.
Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255.
Hirata, S.; Yagi, K. *Chem. Phys. Lett.* **2008**, *464*, 123.
Hirshfeld, F. L.; *Theor. Chem. Acc.* **1977**, *44*, 129.
Hu, H.; Lu, Z. Y.; Yang, W. T. *J. Chem. Theory Comput.* **2007**, *3*, 1004.
Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573.
Hua, D.; Leverentz, H. R.; Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*,

251.
Hua, S.; Hua, W.; Li, S. *J. Phys. Chem. A* **2010**, *114*, 8126.
Hua, S.; Xu, L.; Li, W.; Li, S. *J. Phys. Chem. B* **2011**, *115*, 11462.
Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. *J. Phys. Chem. A* **2008**, *112*, 10864.
Huang, L.; Massa, L.; Karle, J. *J. Int. Quantum Chem.* **2005**, *103*, 808.
Huang, L.; Massa, L.; Karle, J. *J. Int. Quantum Chem.* **2006**, *106*, 447.
Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
Jiang, N.; Ma, J.; Jiang, Y. S. *J. Chem. Phys.* **2006**, *124*, 114112.
Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.
Kairys, V.; Jensen, J. H. *Chem. Phys. Lett.* **1999**, *315*, 140.
Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133.
Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
Kitaura, K.; Sugiki, S. I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* **2001**, *336*, 163.
Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1990**, *172*, 243.
Kohn, W.; Sham, L. *J. Phys. Rev.* **1965**, *140*, A1133.
König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.
Köster, A. M.; Kölle, C.; Jug, K. *J. Chem. Phys.* **1993**, *99*, 1224.
Kumar, R.; Wang, F. F.; Jenness, G. R.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309.
Kurbanov, E.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 1.
Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 7300.
Le, H. A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2009**, *113*, 10527.
Le, H. A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A *J. Chem. Theory Comput.* **2012**, *8*, 469.
Lee, A. M.; Bettens, R. P. A *J. Phys. Chem. A* **2007**, *111*, 5111.
Lee, C.; Yang, W. *J. Chem. Phys.* **1992**, *96*, 2408.
Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573.
Leverentz, H.; Gao, J.; Truhlar, D. G. *Theor. Chem. Acc.* **2011**, *129*, 3.
Lewin, J. L.; Cramer, C. J. *J. Phys. Chem. A* **2008**, *112*, 12754.
Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.
Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.
Li, W.; Piecuch, P. *J. Phys. Chem. A.* **2010**, *114*, 6721.
Li, W. J. *Chem. Phys.* **2013**, *138*, 14106.
Lillestolen, T. C.; Wheatley, R. *J. Chem. Comm.* **2008**, 5909.
Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.
Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
López, N.; Illas, F. *J. Phys. Chem. B* **1998**, *102*, 1430.

Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365.
Lu, J. P.; Yang, W. *Phys. Rev. B* **1994**, *49*, 11421.
Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
Mahadevi, A. S.; Rahalkar, A. P.; Gadre, S. R.; Sastry, G. N. *J. Chem. Phys.* **2010**, *133*, 164308.
Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. Atomic Radii of the Elements. In CRC Handbook of Chemistry and Physics; 92nd; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, **2011-2012**; pp. 9-49f.
Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527.
Martin, R. M. Electronic Structure: Basic Theory and Practical Methods; Cambridge University Press: New York, **2004**.
Maseras, F.; Morokuma, K. *J. Comp. Chem.* **1995**, *16*, 1170.
Massey, H. S. W.; Burhop, E. H. S. The international series of monographs on physics-Electronic and ionic impact phenomena, 2 ed.; University Press: Oxford, 1969; Vol. 1.
Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336.
Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669.
Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592.
Moon, S.; Christiansen, P. A.; DiLabio, G. A. *J. Chem. Phys.* **2004**, *120*, 9080.
Moret, M.-E.; Tapavicza, E.; Guidoni, L.; Röhrig, U.; Sulpizi, M.; Tavernelli, I.; Rothlisberger, U. *Chimia* **2005**, *59*, 493.
Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393.
Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
Murata, K.; Fedorov, D. G.; Nakanishi, I.; Kitaura, K. *J. Phys. Chem. B* **2009**, *113*, 809.
Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869.
Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
Nasluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, *N. J. Phys. Chem. B* **2003**, *107*, 2228.
Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc.* **2001**, *106*, 105.
Nygren, M. A.; Pettersson, L. G. M.; Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1994**, *100*, 2010.
Ohnishi, Y. Y.; Nakao, Y.; Sato, H.; Sakaki, S. *J. Phys. Chem. A* **2008**, *112*, 1946.
Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046.
Pacchioni, G.; Ferrari, A. M. *Catal. Today* **1999**, *50*, 533.
Pacios, L. F.; Christiansen, P. A. *J. Chem. Phys.* **1985**, *82*, 2664.
Parks, J. M.; Hu, H.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 154106.
Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353.
Piquemal, J. P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.

Politzer, P., Truhlar, D. G., Eds.; Chemical Applications of Atomic and Molecular Electrostatic Potentials; Plenum: New York, **1981**.
Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
Praprotnik, M.; Delle Site, L.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545.
Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7752.
Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.
Rahalkar, A. P.; Ganesh, V. ; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101.
Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405.
Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
Redondo, A.; Goddard, W. A.; Swarts, C. A.; McGill, T. C. *J. Vac. Sci. Technol.* **1981**, *19*, 498.
Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066.
Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.
Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736.
Reinhardt, P.; Piquemal, J. P. *Int. J. Quantum Chem.* **2009**, *109*, 3259.
Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
Řezač, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2009**, *6*, 91.
Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458.
Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
Rotenberg, B.; Salanne, M.; Simon, C.; Vuilleumier, R. *Phys. Rev. Lett.* **2010**, *104*, 138301.
Sawada, T.; Fedorov, D. G.; Kitaura, K. *Int. J. Quantum Chem.* **2009**, *109*, 2033.
Senn, H. M.; Thiel, W. *Angew. Chemie Int. Ed.* **2009**, *48*, 1198.
Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182.
Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79.
Sherwood, P. In Modern Methods and Algorithms of Quantum Chemistry, Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, **2000**; pp 285.
Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *J. Mol. Struct. Theochem*. **2003**, *632*, 1.
Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630.
Shurki, A.; Warshel, A. Advances in Protein Chemistry. In Adv.Protein Chem.; Academic Press Inc: San Diego, USA, **2003**; Vol. 66; pp 249.
Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
Slater, J. C. *Phys. Rev.* **1930**, *36*, 57.
Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.

Soave, R.; Pacchioni, G. *Chem. Phys. Lett.* **2000**, *320*, 345.
Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085
Sokalski, W. A.; Poirier, R. A. *Chem. Phys. Lett.* **1983**, *98*, 86.
Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 683.
Sousa, S. F.; Ramos, M. J. In Computational Proteomics 2008; Ramos, M. J., Ed.; Transworld Research Network: Kerala, India, **2008**; pp 101.
Spackman, M. A. *Chem. Phys. Lett.* **2006**, *418*, 158.
Speetzen, E. D.; Leverentz, H. R.; Lin, H.; Truhlar, D. G. Accurate Condensed Phase Electronic Structure Theory; Manby, F., Ed.; CRC Press: Boca Raton, **2010**.
Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
St.-Amant, A.; Cornell, W. D.; Kollman, P. A.; Halgren, T. A. *J. Comput. Chem.* **1995**, *16*, 148
Stewart, R. F. *J. Chem. Phys.* **1969**, *50*, 2485.
Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
Stone, A. J. The Theory of Intermolecular Forces; Oxford University Press Inc.: New York, **1996**.
Stone, A. J. *J. Phys. Chem. A* **2011**, *115*, 7017.
Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.
Strain, M. C.; Scuseria, G. E.; Frisch, N. J. *Science* **1996**, *271*, 51.
Strand, T. G.; Bonham, R. A. *J. Chem. Phys.* **1964**, *40*, 1686.
Suarez, E.; Diaz, N.; Suarez, D. *J. Chem. Theory Comput.* **2009**, *5*, 1667.
Szabo, A.; Ostlund, N. S. Modern Quantum Chemistry-Introduction to Advanced Electronic Structure Theory; Dover Publications, Inc.: Mineola, New York, **1989**.
Tafipolsky, M.; Amirjalayer, S.; Schmid, R. *Micropor. Mesopor. Mater.* **2010**, *129*, 304.
Tafipolsky, M.; Engels, B. *J. Chem. Theory Comput.* **2011**, *7*, 1791.
Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.
Torheyden, M.; Jansen, G. *Mol. Phys.* **2006**, *104*, 2101.
Truhlar, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 16824.
Valero, R.; Gomes, J. R. B.; Truhlar, D. G.; Illas, F. *J. Chem. Phys.* **2008**, *129*, 124710.
von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.
Vreven, T.; Morokuma, K.; David, C. S. *Annu. Rep. in Comput. Chem.* **2006**, *2*, 35.
Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.
Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330.
Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 359.
Wang, B.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10556.
Wang, Y.; Sosa, C. P.; Cembran, A.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781.
Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
Weber, J.; Gunne, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 583.

Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578.
Wen. S.; Nanda, K.; Huang, Y.; Beran, G. *Phys. Chem. Chem. Phys*. **2012**, *14*, 7578.
Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. *J. Phys. Chem. A* **2007**, *112*, 268.
Wheatley, R. J.; Mitchell, J. B. O. *J. Comput. Chem.* **1994**, *15*, 1187.
White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.
White. C. A. ; Head-Gordon, M. *J. Chem. Phys.* **1994**, *101*, 6593.
Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys*. **2008**, *128*, 234108.
Yang, W. *Phys. Rev. A* **1991**, *44*, 7823.
Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674.
Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2010**, *132*, 094102.
Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2011**, *134*, 084111.
York, D.; Lu, J. P.; Yang, W. *Phys. Rev. B* **1994**, *49*, 8526.
Yudanov, I. V.; Nasluzov, V. A.; Neyman, K. M.; Rösch, N. *Int. J. Quantum Chem.* **1997**, *65*, 975.
Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
Zhang, Y. *Theor. Chem. Acc.* **2006**, *116*, 43.
Zhao, Q.; Yang, W. *J. Chem. Phys*. **1995**, *102*, 9598.
Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.
Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.